

THIS WEEK

EDITORIALS

WORLD VIEW The first steps on the slippery slope to academic fraud **p.151**

GRAVITY GRAVE Black-hole survey puts down minimum size marker **p.152**

FOSSIL Old bone shows ancient beaked whales had skeletal worms **p.153**



Academic freedom

A court decision in the United States rescinding an order to turn over academic e-mails in response to a freedom-of-information request is welcome.

Scottish law exempts academic work from the freedom-of-information laws, but the rest of the United Kingdom does not. Ireland also exempts, and although the United States is commonly thought to, it turns out that, as so often in that country, it is left to the courts to decide. So, just what should researchers make of freedom-of-information laws?

American climatologist Michael Mann, now at Pennsylvania State University in University Park, probably knows the score better than most. And in the latest twist in a long-running saga over who should be able to read Mann's e-mails, a Prince William County Circuit Court judge in Manassas, Virginia, last week tore up an agreement that would have given the e-mails, with conditions, to attorneys for the American Tradition Institute (ATI), a conservative think tank. Judge Gaylord Finch also granted Mann's request to join the University of Virginia (his former employer and holder of the e-mails) in a lawsuit to block their release.

As both sides argue about whether the messages should ultimately be made public, the two legal decisions come as welcome news to those (including this journal) who believe that access to personal correspondence is a freedom too far. But the case highlights, yet again, how woefully unprepared the academic community is to meet this kind of challenge. This must change.

Certainly, the University of Virginia caved in too easily when it signed the agreement that granted the ATI access to the e-mails last spring. Given the tone of public statements that have come out of the ATI, the university should never have agreed to hand over confidential material of any sort.

But the university and its attorneys deserve credit for rectifying the situation. And despite appearances, to fight such requests is not against the letter, or indeed the spirit, of perfectly proper regulations designed to improve the accountability of public bodies. In fact, Virginia's freedom-of-information law provides the university with a solid basis to deny access to this kind of blanket request for e-mail records: academic work is exempt. This is as it should be, and the university should fight to protect that exemption now and in the future.

Yes, the public has a right to know, and yes, greater scrutiny of public spending is a good thing. But research practice is typically protected for good reasons too. To protect academic freedom is a foundation for intellectual property and copyright laws, while in court, both Mann and the university warned of the chilling effect of such demands on communication between scientists. Certainly, many researchers are more wary of e-mail today, and given Mann's experiences, who can blame them?

His case is high profile, but scientists and academics watching it (as well as the related attempts by Virginia's attorney-general Ken Cuccinelli to force the release of the same e-mails) should be cautious about drawing broad conclusions from how it may pan out. Even within the United States, the eventual ruling won't serve as much of a

precedent outside Virginia. Federal agencies in the United States are subject to the federal statute, but state universities and research institutions must all play by the laws enacted in their own states.

Across those states it seems that this kind of academic exemption is common, but not universal, and its application would vary according to precedents set locally. In other words, it will be up to individual universities to work out how to address these kinds of cases as they emerge in future.

Mann's decision to join the lawsuit was spurred by the initial decision of the university to grant ATI access to his e-mails, a move with which he disagreed. He suggests that universities may be limited in what they can do to fend off these attacks, or that their interests may not always align precisely with those of individual researchers.

Mann is also getting help from a new fund especially designed to aid climate scientists hit by legal challenges, and organizations including the American Geophysical Union, the American Association of University Professors and the Union of Concerned Scientists have weighed in as well. All of this is good and useful, but it is no substitute for a solid institutional defence. Individual universities and research institutions everywhere should review their own policies and make sure they know the applicable laws as well as do those who would use them for mischief, or worse. ■

Innovative vision

Bill Gates gave the G20 summit a workable plan to boost development around the world.

What a shame that the latest lurch of the financial crisis in Greece and the eurozone overshadowed all else at last week's G20 summit in Cannes, France. For on the agenda was a brief but important report on ways to boost funding for development, research and innovation in health and agriculture. If implemented, its suggestions would stimulate innovation and go a long way towards helping to alleviate poverty, hunger and disease. The report came from computer-tycoon-turned-philanthropist Bill Gates, and although the typically vague final G20 communiqué gave his suggestions only brief mention, that they feature at all in the current climate is a notable achievement.

Gates, the first private individual to address a G20 summit, pleaded for countries not to let the financial crisis cause them to renege on their existing pledges, which would generate an additional US\$80 billion annually from 2015 onwards.

Some programmes, such as the GAVI Alliance — formerly the Global Alliance for Vaccines and Immunisation — have seen pledges increase this year (see go.nature.com/qlldf4), and donors must follow through with the cash. Others have not been so lucky. Donations to the Global Fund to Fight AIDS, Tuberculosis and Malaria, for example, last year fell far short of its funding target (see *Nature* **467**, 767; 2010). And the comprehensive Global Funding of Innovation for Neglected Diseases (G-Finder) report, due next month, is expected to say that most donors slashed funding for neglected-disease research and development last year — causing an overall fall of more than \$100 million, or more than 5%. The risk is that the financial crisis could roll back the huge progress that has been made in both funding and outcomes for global health and research since neglected diseases returned to the international agenda in the mid-1990s — and also stymie a recent resurgence of interest in agricultural research and development (R&D) for developing nations.

To combat this threat, Gates rightly emphasized the urgent need for new funding mechanisms to boost development and make it less vulnerable to financial turmoil. And he made a compelling case for measures that, between them, could potentially raise more than \$100 billion a year. Gates also put his finger on a key point: cash flow between rich and poor countries is not a one-way street of aid from donors to recipients. Many poorer nations have substantial natural resources, the revenues from which exceed that of aid. Yet countries are sometimes given raw deals by foreign companies exploiting those resources, and revenues can also end up in the bank accounts of corrupt public officials. The result is a hemorrhaging of financial resources, some of which could otherwise be spent on building labs, hospitals and sanitation systems, training researchers and doctors, or buying bed nets and drugs.

To tackle this, Gates called on the G20 countries to embrace the Extractive Industries Transparency Initiative (EITI), a World Bank-backed scheme, launched in 2002, to oblige companies and countries to make public the terms of oil, gas and mineral deals in order to better

monitor both whether the deals are fair and where that money goes. The sums involved are potential game changers that could also transform neglected diseases and agricultural R&D. Gates points out that, at peak production, Uganda's oil reserves are estimated to generate \$2 billion annually, which is almost as much as the country's entire national budget of \$3 billion. However, the confidentiality of the terms of deals with firms makes it impossible to track either whether countries are getting good deals, or where all the cash goes.

“Cash flow between rich and poor countries is not a one-way street.”

The EITI is gaining traction, and teeth, with US President Barack Obama announcing in September that the United States would adopt legislation to make it EITI compliant, and the European Union is considering following suit. But Gates is right to call on all G20 countries to endorse the EITI, and to force companies listed on their stock exchanges to disclose the royalties they pay to governments — and for that measure to be extended to resources such as land and timber.

Gates also called for a share of sovereign wealth funds to be invested in infrastructure, and lent his support to proposals for a small tax on tobacco and financial transactions, and a carbon tax on aviation and shipping fuel, which together could raise at least tens of billions of dollars. Financial-transaction taxes already exist in several countries, and, as Gates said, “are clearly technically feasible”. Likewise, UNITAID, an international organization that helps to accelerate development and scale-up of access to treatments for HIV and AIDS, malaria and tuberculosis, is largely financed by an airline tax paid by its member states.

Gates deserves great credit for highlighting these issues and helping to keep them on the international agenda. Research leaders and politicians must press for them to remain there, and for action to follow. It would be a fitting result if the man whose operating systems forced the world to learn the keyboard sequence CTRL-ALT-DEL can spark a much-needed reboot of funding of research for development. ■

Hubble cleared

A painstaking study absolves US astronomer Edwin Hubble of censoring a Belgian rival.

Edwin Hubble is that relatively rare thing among dead astronomers — a global household name. He owes his status mainly to the NASA space telescope named in his honour. So when researchers suggested this year that Hubble might have censored the work of a rival to secure credit for the groundbreaking discovery that the Universe is expanding, they triggered a fuss that was far removed from the usual arcane wrangling over historical research priority.

In an admirably thorough Comment on page 171, Mario Livio, an astronomer at the Space Telescope Science Institute in Baltimore, Maryland, clears Hubble of wrongdoing. As a result, NASA and a generation of researchers whose careers are closely tied to the Hubble brand can look skywards with some relief.

The charges against Hubble certainly warranted examination. In 1927, the Belgian astronomer Georges Lemaître published a French-language paper in the *Annales de la Société Scientifique de Bruxelles* that laid out the essentials of a picture of galaxies expanding away from one another, and derived an expansion parameter on the basis of then-recent observations. In 1929, Hubble independently put forward and confirmed the same idea, and the parameter later became known as the Hubble constant. In 1931, Lemaître's paper was translated into English and published in the *Monthly Notices of the Royal Astronomical Society*, but most English speakers probably learned of Hubble's contribution before they learned of Lemaître's.

Suspensions of foul play emerged earlier this year, when amateur

historians noticed that the derivation of the expansion constant is missing from the English translation of Lemaître's work. Knowing that Hubble was concerned that he, and the Mount Wilson Observatory in Pasadena, California, at which he made his observations, should get ample credit for confirming the expansion of the Universe, it was tempting to speculate that he had a hand in the editing of the Belgian's paper. But motive alone doesn't build a case, and professional historians, who had known of the irregularity for years, remained sceptical.

Livio's research suggests that they were right to hesitate. After reviewing hundreds of documents in the archives of the Royal Astronomical Society in London, Livio found a copy of a 1931 letter by Lemaître in which he said that in translating his paper, he had deleted discussion of the velocities of galaxies because it was “of no actual interest”. Why exactly Lemaître thought this is unclear, but it seems that he was not very concerned about getting the credit for his work in the way that modern followers have assumed; instead, he may have worried more about seeming out of date, given that the data on which the expansion constant was based had been improved since 1927.

The idea that the accuracy of papers and their relevance to colleagues ought to be more important than ensuring priority at every step may seem fantastic in today's cut-throat world of science. And perhaps it was then, too. Perhaps Lemaître was simply so flattered to be invited to translate his paper that, aware of Hubble's importance among English-speakers and fearful of repercussions, or eager to join the Royal Astronomical Society, he self-censored. The case against Hubble is closed, but there may still be a story for motivated historians to look into.

Space agencies should also take note. Whether or not Hubble deliberately censored Lemaître, the fact is that in the English-speaking world, Lemaître has lost — to Hubble — priority for his contributions. The Belgian's name is a worthy candidate for the title of a future space mission. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunjv



The road to fraud starts with a single step

The extensive academic fraud of Diederik Stapel has rocked science. Social psychologist Jennifer Crocker traces the destructive path that cheats follow.

Fraud happens uncomfortably often — from financier Bernie Madoff to the (now imprisoned) real-estate developer who built the outsized house, recently foreclosed, in my neighbourhood. But it disturbs most when it happens close to our professional home.

Diederik Stapel, a social psychologist and author of many published papers, has resigned his position at Tilburg University in the Netherlands after admitting to fabricating data in his research (see *Nature* 479, 15; 2011). I know Stapel through limited interactions at conferences, and he won the Early Career Award from the International Society for Self and Identity when I was its president in 2007.

He published his findings in several journals of the American Psychological Association, for which I am chair of the Publications and Communications Board. The association, and other publishers, will retract any fraudulent works that the investigation identifies. The early signs are that the scale of his fraud is vast.

Such cases of outright fraud in science are distressing for many reasons. For example, they damage the careers of students and collaborators, and raise doubts about all papers by the same author. Most importantly, they damage public trust in science and in scientists. In this case, trust in social psychologists, and the work we do, has been undermined.

How can it happen? Why would someone with obvious intelligence, ambition and talent risk everything by falsifying data? Social psychology offers us a way to answer such questions.

To understand fraud, we should think about how it begins and escalates, not how it ends. By the time such fraud is exposed, bad choices that would usually lead to only minor transgressions have escalated into outright career-killing behaviour.

Stanley Milgram's 1963 classic studies of obedience to authority help to show why people do things that are highly counter to their norms. Milgram's studies are usually interpreted as providing insight into how situations determine behaviour. Another lesson of his experiments is how easy it is to take the first small step on the slippery slope that ends with violation of our norms and values, and how hard it is to stop.

In Milgram's study, research participants were told to administer escalating electric shocks when actors answered questions incorrectly. The work is famous because of where it ended. More than half the subjects were willing to administer shocks beyond the point at which the 'learner' complained of heart pain and then stopped responding. The lesson of the study seemed to be that people would violate their own moral codes and administer potentially deadly shocks on the say-so of an authority figure.

But look at where the experiment begins.

Every one of the participants started by giving only a 'slight' shock of 15 volts in response to the learner's first incorrect answer. With assurances from the supervising experimenter that the shock might be painful but was not dangerous, what could be the harm?

The harm is that once people have delivered 15 volts of shock, they have no compelling reason to resist 15-volt increases. After all, they have implicitly conceded that 15 volts of shock is minor. Each time participants administered an increased shock, that level of shock became the new normal. Consciously or unconsciously, they justified their behaviour to themselves every time they pulled the switch, and every justification made the next pulling of the switch easier.

To understand fraud in science, the useful lesson is the significance of that first tiny step. Every minor transgression — dropping an inconvenient data point, or failing to give credit where it is due — creates a threat to self-image. The perpetrators are forced to ask themselves: am I really that sort of person? Then, to avoid the discomfort of this threat, they rationalize and justify their way out, until their behaviour feels comfortable and right. This makes the next transgression seem not only easier, but even morally correct.

The well-being of science and our society requires that fraud be punished severely. But a heavy focus on fraudsters may also conveniently divert our attention from the fraudster within us all. Who cannot find places where they took a first step, or perhaps several steps, down one slippery slope or another? The road to fraud probably starts out with a step taken because of some egotistical fear or anxiety — fear of losing someone's respect, for example, or of letting others

down, the fear of being seen as a loser, of being a failure, or of not getting the job, the grant or the award that one covets.

In such circumstances, the difficult question then becomes, how can we stop the slide? In the case of the '15-volt' steps towards scientific misconduct, thinking about the consequences for our students, colleagues, loved ones, our institution, our discipline or science itself might halt our own little slides. In this regard, we should all admire the colleagues and researchers who took the risk to stop something unacceptable when they saw it in the Stapel case. Surely, they too experienced egotistical fears. Will people believe me? What will happen to me? Will my own reputation be tarnished?

The slippery slope beckoned, but they acted for the common good, and we should thank them. ■

Jennifer Crocker is a social psychologist at Ohio State University and past president of the Society for Personality and Social Psychology (SPSP). e-mail: crocker.37@osu.edu

This article is adapted from a post on the SPSP website.

A HEAVY FOCUS ON
FRAUDSTERS
MAY ALSO
CONVENIENTLY
DIVERT OUR
ATTENTION FROM THE
FRAUDSTER WITHIN
US ALL.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/bprcvl

SEVEN DAYS

The news in brief

RESEARCH

Fracking tremors

A British energy company says that its hydraulic fracturing ('fracking') project probably caused the cluster of small earthquakes that struck Lancashire, UK, this spring. Cuadrilla Resources, based in Lichfield, had commissioned independent studies to investigate the quakes, the largest of which was of magnitude 2.3. The 2 November synthesis report says that it should be safe to continue operations, although protest groups disagree. Concerns have been raised in many countries about the safety of fracking, a technique in which high-pressure fluids are pumped into shale to fracture the rock and force out natural gas. See go.nature.com/p5fj1q for more.

EVENTS

Heavenly kiss

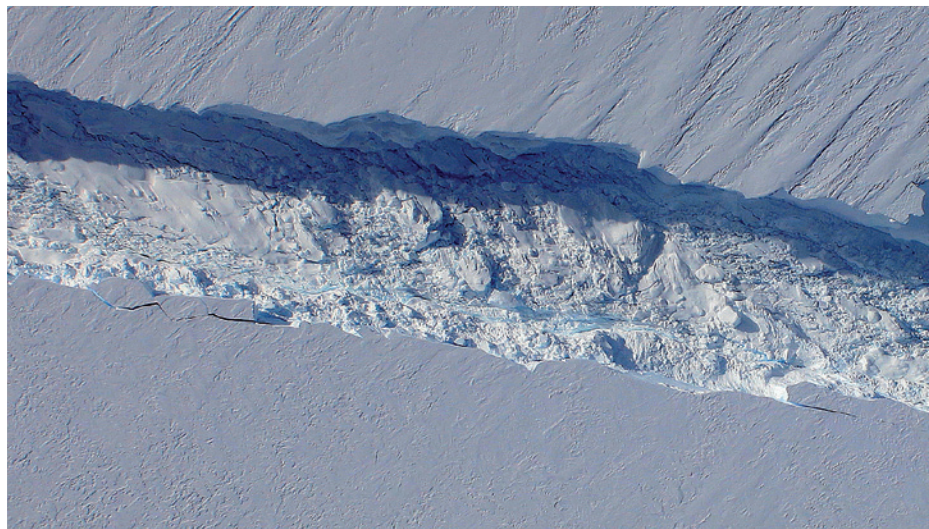
China's unmanned Shenzhou 8 spacecraft docked with its Tiangong 1 module on 2 November, marking the country's first success at the delicate procedure, which was broadcast live on state television. After completing what state media referred to as a 'kiss' in space, both craft are now orbiting Earth together. Two more missions, perhaps

NUMBER CRUNCH

\$131 bn

Lifetime sales of Pfizer's cholesterol-lowering drug Lipitor (atorvastatin), the world's top-selling prescription medicine. It loses patent protection from 30 November.

Source: Reuters, IMS Health



NASA/DMS

How icebergs begin

A seasonal ice-survey flight has spotted the birth signs of a large iceberg in West Antarctica: a crack at least 18 miles long in Pine Island Glacier, which sticks out from the Amundsen Sea coast on the west of the continent. The rift (pictured, around 80 metres wide and 50 metres deep) was first seen in mid-October

by NASA's Operation IceBridge project, which released images last week. The iceberg that eventually calves from the breaking ice shelf will cover around 880 square kilometres, project scientists said. Pine Island Glacier is rapidly retreating, accounting for a large part of West Antarctica's ice loss.

carrying astronauts, will follow in 2012. If the next stages of testing go to plan, China will launch further modules to be assembled into a space station by 2020.

Back from 'Mars'

Six men have survived 520 days cooped up in 3 small rooms at the Institute of Biomedical Problems in Moscow, where they were simulating the isolation of a journey to Mars and back. On 4 November, the crewmen — three of whom were Russian, one Chinese, one French and one Italian — emerged pale but smiling from the capsule they had entered in June 2010. Many observers found fault with the exercise's lack of realism, saying that the mission could not simulate

weightlessness, true danger or the pressure and motivation of a real journey to Mars. See go.nature.com/1zquiu for more.

Nuclear restart

For the first time since the Fukushima disaster, a Japanese nuclear reactor that had gone into shutdown has been brought back online. The Genkai Nuclear Power Plant in southwestern Japan was shut down briefly in October because of a technical fault, and its restart last week would normally be unremarkable. But concerns over nuclear safety mean that no other plants closed since the earthquake and tsunami on 11 March have been allowed to restart. Of Japan's 54 nuclear reactors, 43 are currently offline.

Asteroid fly-by

Marshalling everything from major radar facilities to backyard telescopes, astronomers geared up this week for a fantastic view of an asteroid called 2005 YU55. The 400-metre-diameter rock passed Earth on 8 November just 320,000 kilometres away, or 0.85 of the distance between Earth and the Moon. It is the closest pass by an asteroid this big since 1976, and there won't be another until 2028. See go.nature.com/ohn6zt for more.

BUSINESS

Avandia fine

Drug giant GlaxoSmithKline (GSK) announced on 3 November that it has agreed to pay US\$3 billion

R. HAHN/FERMILAB
to settle a bevy of US federal investigations into the way it has developed and marketed some of its biggest-selling drugs. Foremost among these is Avandia (rosiglitazone), a once-dominant diabetes drug, sales of which were banned in Europe and restricted in the United States last year after concerns that it increased risks of heart attack and stroke. The company, headquartered in London, came under fire in July 2010, when a US Senate committee concluded that GSK had known about the drug's heart risks for more than a decade without reporting them to regulators. See go.nature.com/iphvp for more.

PEOPLE

SLAC head leaves

Persis Drell is stepping down as the director of the SLAC National Accelerator Laboratory in Menlo Park, California; the lab's operator, Stanford University, announced the decision on 1 November. Drell became SLAC's fourth director in 2007, succeeding Jonathan Dorfan. She will stay on until a replacement is found, and then return to research at Stanford.

Nobel physicist dies

Atomic physicist Norman Ramsey, who shared the 1989 Nobel Prize in Physics,



died on 4 November, aged 96. After working on radar and the atomic bomb in the Second World War, Ramsey (**pictured**) moved to Harvard University in Cambridge, Massachusetts, where he developed improved methods for probing atomic structure by measuring the response to electromagnetic radiation. His discoveries, which won him the Nobel prize, led to the caesium atomic clock and the hydrogen maser. Ramsey also helped to found Brookhaven National Laboratory in Upton, New York, and Fermilab in Batavia, Illinois.

POLICY

Drug approvals rise

The US Food and Drug Administration approved more innovative drugs this year than in any year of the past decade, except for 2009, the agency said last week. In the US government's

fiscal year 2011, which ended on 30 September, the agency approved 35 original medicines, 24 of which were authorized in the United States before anywhere else. Ten were for rare, or 'orphan', diseases and two were 'personalized' medicines: treatments for melanoma and lung cancer that were approved along with diagnostic tests to identify the patients that they are most likely to help.

Public integrity

More US agency policies on scientific integrity seem likely to become public after John Holdren — director of the White House Office of Science and Technology Policy (OSTP) — encouraged officials to release the tightly guarded guidelines. In a 31 October letter, Holdren said he had set a 17 December deadline for agencies to submit revised policies that take account of OSTP feedback on earlier drafts. Nineteen agencies have policies in the works, but so far only six are publicly available. Critics say that the policies lack provisions to protect whistleblowers from retaliation.

Future medicine

The US National Research Council (NRC) has called for a network that would connect patients' health records with

COMING UP

12–16 NOVEMBER

The Society for Neuroscience holds the world's largest annual neuroscience meeting in Washington DC. A featured panel looks at the interplay between economics and the brain. www.sfn.org/am2011/

16 NOVEMBER

The Commission on Sustainable Agriculture and Climate Change — a global task force of senior scientists and government advisers — presents recommendations for avoiding future food crises. ccafs.cgiar.org/commission

layers of data on molecular tests, genetics, and social and physical environments. This, the NRC said in a report published on 2 November, would enable treatments to be personalized for patients, ushering in a new taxonomy of human disease based on molecular origins rather than on physical signs and symptoms. The report was requested last year by the National Institutes of Health. See go.nature.com/ayesl2 for more.

Carbon tax

Australia will introduce a tax on carbon dioxide emissions from 1 July 2012, after the legislation passed its final barrier — approval by the Senate — on 8 November. The bill had been passed by the House of Representatives in October. The tax, of Aus\$23 (US\$24) per tonne for the country's top 500 emitters, will increase by 2.5% a year above inflation until an emissions-trading scheme replaces it in 2015.

NATURE.COM

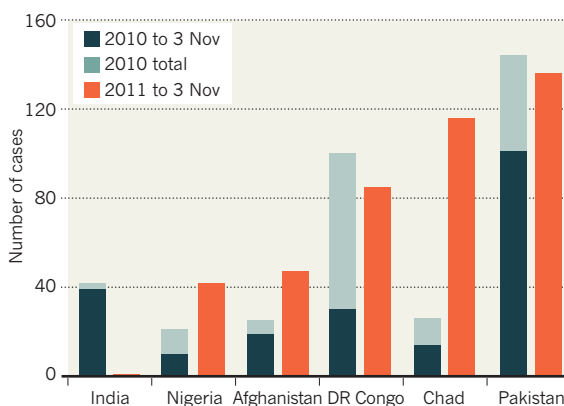
For daily news updates see: www.nature.com/news

TREND WATCH

Efforts to eradicate polio are bearing fruit in India, one of four countries (with Nigeria, Pakistan and Afghanistan) where wild poliovirus (WPV) is endemic. India reported only one confirmed WPV case this year, in January; the absence of the virus, particularly during India's high-transmission season between June and November, is unprecedented. But India is a lone bright spot: other countries in which polio still persists have already had more cases than this time last year (see chart).

POLIO CLINGS ON

India seems on track to stop poliovirus transmission this year, but the disease is springing back in other countries.



NEWS IN FOCUS

PHYSICS Leap second living on borrowed time **p.158**

ENVIRONMENT The great Amazon hydropower rush **p.160**

LAW Thalidomide tragedy spawns fresh round of lawsuits **p.161**

PROFILE Seth Stein scoffs at mid-continent quake fears **p.166**



L. MACGREGOR/REUTERS



Andrew Wakefield's discredited theory linking vaccination and autism stirred public fears.

AUTISM

Fresh dispute about MMR 'fraud'

Pathology records are at the centre of a new disagreement over disgraced medic Andrew Wakefield.

BY EUGENIE SAMUEL REICH

It is one of the most serious allegations that could be made about a doctor: manipulating patients' histories to make money. So it is no wonder that the charges, levied by editors of the *British Medical Journal* (BMJ) in January against medical researcher Andrew Wakefield, are still getting close scrutiny. Now an American whistleblower advocacy group has joined

the fray over Wakefield, who in 1998 hypothesized a link, now scientifically disproven, between the measles, mumps and rubella vaccine (MMR) and autism.

On 9 November, David Lewis of the National Whistleblower's Center in Washington DC published a letter in the *BMJ* (bmj.com) arguing that Wakefield did not commit research fraud. Lewis told *Nature* that he thinks the combination of public charges

and a slow, secretive investigation has left the public not knowing whom to believe and is unfair to the accused researcher. "[The system] throws people like Andy into a no-man's-land," Lewis says.

Wakefield was the lead author of a 1998 paper in *The Lancet* (A. J. Wakefield *et al. Lancet* 351, 637–641; 1998) reporting on the case histories of 12 children who had received the MMR vaccine and developed symptoms of autism or inflammatory bowel disease.

The paper inflamed public fears about vaccines, but it was retracted in 2010 after the UK General Medical Council (GMC) concluded that Wakefield had a charge of serious professional misconduct to answer, in part because it found that his team did not have proper ethical approval for tests performed on the children. Later in the year, the GMC found him guilty of the misconduct charge and revoked his licence to practice as a doctor. By then, more than 12 large-scale epidemiological studies had failed to find evidence of the hypothesized link (J. S. Gerber and P. A. Offit *Clin. Infect. Dis.* 48, 456–461; 2009) and the MMR vaccine is today regarded as safe.

But was the paper merely mistaken, or a deliberate fraud? Articles by medical journalist Brian Deer published in the *BMJ* in 2010 and 2011 accused Wakefield of reporting histories for the children that were not consistent with their records and their parents' recollections, at a time when Wakefield was also being paid by lawyers intending to sue MMR manufacturers. Deer's articles themselves did not allege fraud, but on their basis a *BMJ* editorial in January 2011 called the paper fraudulent.

Wakefield has always denied allegations of manipulating data or having a financial motive in the *Lancet* paper, but no institution has yet ruled on the matter. In March, University College London (UCL), which took over divisions of the Royal Free Hospital where Wakefield worked, said it had launched an investigation.

In the meantime, Lewis, a microbiologist and former whistleblower who says he was falsely accused of misconduct after alleging links between human illness and the spread-

ing of sewage sludge, asked Wakefield for the chance to review his files. His claims are not likely to challenge the conclusion — based on much

➔ **NATURE.COM**
For more on
unfounded vaccine
fears see:
go.nature.com/a3nqzx

► other evidence — that the vaccine does not trigger autism. But they could complicate the debates about Wakefield's integrity and the UCL investigation.

The documents that Lewis reviewed include confidential forms describing biopsies from the guts of children. The forms were filled out by pathologists Andrew Anthony and Paul Dhillon, who worked with Wakefield at the Royal Free. These documents, Lewis says, are relevant to Deer's charge that records he obtained do not support Wakefield's claims in the *Lancet* paper that the children had nonspecific colitis, a supposed element of an MMR-induced syndrome. On sheets for three of the children graded by Anthony, the handwritten word "colitis" appears, and Dhillon checked a box labelled "non-specific" on 10 forms. Anthony's sheets are dated after the *Lancet* publication, whereas Dhillon's are dated before.

Lewis believes that the sheets show that Anthony and Dhillon were making good-faith

diagnoses of colitis. Anthony, who has left UCL, could not be reached by *Nature*, and Dhillon indicated that UCL had told him not to comment. (Neither has been accused of manipulating data.)

Before publishing Lewis's letter, the *BMJ* asked Ingvar Bjarnason, a gastroenterologist at King's College Hospital, London, to review the materials. Bjarnason says he doesn't believe they are sufficient to support claims in the *Lancet* paper of a new disease process. He also questions whether "non-specific" on the grading sheets refers to colitis, saying it could refer to any kind of gut changes. But he says that the forms don't clearly support charges that Wakefield deliberately misinterpreted the records. "The data are subjective. It's different to say it's deliberate falsification," he says.

Deer notes that he never accused Wakefield of fraud over his interpretation of pathology records. But he says that records read to him from the Royal Free pathology service

clearly stated that the children's gut biopsies were within normal limits, even though they were reported in the *Lancet* paper as having enterocolitis.

Fiona Godlee, the editor of the *BMJ*, says that the journal's conclusion of fraud was not based on the pathology but on a number of discrepancies between the children's records and the claims in the *Lancet* paper. She says she will be calling for a public inquiry into the matter, noting that it has been more than a year since she first informed UCL about concerns over Wakefield's work.

Wakefield denies charges of data manipulation. He says that UCL has yet to officially request his response to any charges and he isn't convinced that the inquiry will give him a fair hearing. A UCL spokesman says that the investigation will be "thorough, fair and wide-ranging". But eight months after announcing its inquiry, the university is still looking for a suitable external chairman. ■

TECHNOLOGY

Time is running out for the leap second

Abolition would see 'official' time unmoored from the Sun.

BY ZEEYA MERALI

"The times," sang Bob Dylan, "they are a-changin'." His words could become literal truth in January, when the World Radiocommunication Conference of the International Telecommunication Union in Geneva, Switzerland, will vote on whether to redefine Coordinated Universal Time (UTC) and pull our clock time out of synchronization with the Sun's location in the sky.

At issue is whether to abolish the 'leap second' — the extra second added every year or so to keep UTC in step with Earth's slightly unpredictable orbit. UTC — the reference against which international time zones are set — is calculated by averaging signals from around 400 atomic clocks, with leap seconds added to stop UTC drifting away from solar time at a rate of about one minute every 90 years.

But "leap seconds are a nuisance", says Elisa Felicitas Arias, the director of the Time Department at the International Bureau of Weights and Measures (BIPM) in Sèvres, France. They cannot be preprogrammed into software because they are typically announced only six months in advance by the International Earth Rotation and Reference Systems Service in Frankfurt, Germany. If the seconds

get implemented inconsistently in different systems, clocks can briefly go out of synch, potentially leading to glitches that can stall computers and leave international financial markets vulnerable to attack.

Still, some countries — principally China, Canada and the United Kingdom — want to keep leap seconds to maintain the link with solar time, in part for philosophical reasons. "Most Chinese scholars think it is important for timekeeping to have a connection to astronomical time because of traditional Chinese culture," says Chunhao Han of the Beijing Global Information Center of Application and Exploration, who adds, however, that China has yet to decide how it will vote in January.

Last week, scientists and government representatives met at the Kavli Royal Society International Centre near Milton Keynes, UK, to discuss the issue, but they failed to reach a consensus, making the outcome of the January vote hard to predict. Arias, who co-organized that meeting, argues that leap seconds are obsolete now that global navigation systems, which set their own internal timescales, have replaced solar time for navigation and precision scientific measurements such as the motion of tectonic plates and how Earth's mass warps space-time.

Adding an extra second inconsistently to

multiple clocks across satellite networks could cause a system to fail for long enough to cause an air disaster, says Włodzimierz Lewandowski, a physicist at the BIPM. The US Global Positioning System ignores leap seconds for just this reason, and Russia's GLONASS system has had problems in the past incorporating the leap. Europe's Galileo system, which launched its first two satellites last month, and China's developing BeiDou system will also mark time with their own internal clocks.

But Markus Kuhn, a computer scientist at the University of Cambridge, UK, says that most problems could be overcome by having a consistent prescription for adding extra seconds. Linux operating systems, for example, have experienced problems because they add the whole second in one abrupt jump at midnight, which confuses the software. In September, Google announced that it would use an alternative 'soft-leap' strategy, in which operating systems add portions of the second smoothly

over an extended period. "This should be the standard approach," says Kuhn.

"There's no convincing evidence that anything serious would happen if you made a mistake introducing a leap second."

Peter Whibberley, a physicist at the National Physical Laboratory in Teddington, UK, says that despite ten years of debate, "there's no convincing evidence that anything serious would happen if you made a mistake introducing a leap second into a system". Abolishing leap seconds only defers any problems, he adds. "A century down the line, we'll need to introduce a 'leap minute', and nobody has any sensible arguments for why that won't be a worse issue." ■

POLITICS

Palestinian membership puts UN projects at risk

Science and culture agency loses US funding, and other bodies could follow.

BY DANIEL CRESSEY AND DECLAN BUTLER

Concern is growing that United Nations bodies could face budget crises after the UN agency responsible for science suffered a drastic cut in contributions following a vote in which it admitted the Palestinian Authority as a member.

The vote on 31 October by the UN Educational, Scientific and Cultural Organization (UNESCO), based in Paris, caused the United States to pull its 22% share of the agency's budget — some US\$80 million a year. US law prevents the nation from funding any UN body that grants full membership to any organization or group that is not an internationally recognized state.

Following the vote, Irina Bokova, UNESCO's director-general, said that she was "concerned by the potential challenges that may arise to the universality and financial stability of the Organization". Other countries, such as Canada and Israel, have followed the United States in withdrawing funding; the cuts will put pressure on UNESCO-funded science and education projects.

The US Department of State has warned of a potential 'cascade' effect. As part of continuing attempts to win recognition as an independent state, the Palestinian Authority is reportedly set to apply to join other UN agencies, including the World Health Organization (WHO) in Geneva, Switzerland; the International Atomic Energy Agency (IAEA) in Vienna; and the Food and Agriculture Organization (FAO) in Rome. All of these groups get around one-fifth of their direct funding from the United States, and would lose it if they admitted a Palestinian state.

At UNESCO, 107 nations voted for Palestinian membership and only 14 against, with 52 abstentions. A similar selection of states are members of other UN bodies, so further Palestinian applications have a good chance of success. However, other bodies have different systems for granting membership.

UNESCO funds and runs many international science projects, including a tsunami-warning system in the Indian Ocean, put in place after the 2004 disaster. It also runs science-policy

hagen and a member of the Danish National Commission for UNESCO, says that UNESCO programmes are already "systematically" short of funding. "With a very significant reduction of the budget I fear that significant activities will have to be discontinued entirely and some very hard choices will have to be made," he says.

As *Nature* went to press, the FAO, the WHO and the IAEA said that they had not received applications from the Palestinian Authority. The FAO and the IAEA would not comment on whether they have contingency plans to deal with a potential loss of US funding. A spokesperson for the WHO said that because it has not yet received a membership request, the organization has not reached the point of considering revised funding plans.

Kelley Lee, director of global health at Simon Fraser University in Burnaby, Canada, said that other countries might make up any shortfall in funding if the UNESCO situation was repeated at the WHO. "China is an obvious country to step in," she says.

But a Palestinian membership attempt could be particularly problematic at the WHO. The agency is preoccupied with a huge internal reform driven by its director-general, Margaret Chan (see *Nature* 473, 430–431; 2011), so its members may resist Palestinian overtures. "Overall, my feeling is that it is unlikely to happen given the current and indeed long-standing financial austerity within the organization and Margaret Chan's focus on protecting the organization's financial and political standing," says Lee. "Palestinian membership poses higher risks than I think the organization can bear at present." ■



UN agencies risk their budgets by voting for Palestinian membership.

and education programmes. "The present cuts will seriously reduce the good work currently in hand," says Peter Fensham, emeritus professor of science education at Monash University in Melbourne, Australia, who has worked with UNESCO. He adds, "I also remember how handicapped UNESCO's very good work in science education was when the United States and United Kingdom cut off funding in the Reagan and Thatcher eras."

Jens Jørgen Gaardhøje, a physicist at the Niels Bohr Institute at the University of Copen-

MORE ONLINE



TOP STORY

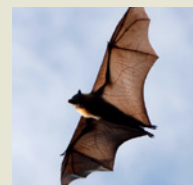
Nobel laureates age like a fine wine
go.nature.com/vwoidq

MORE NEWS

- Tyrannosaurs were power-walkers
go.nature.com/rvjcsw
- Russia takes aim at Phobos
go.nature.com/6yapyx
- Nanoparticle solar cells offer fresh promise for printable photo voltaics
go.nature.com/ohv2eq

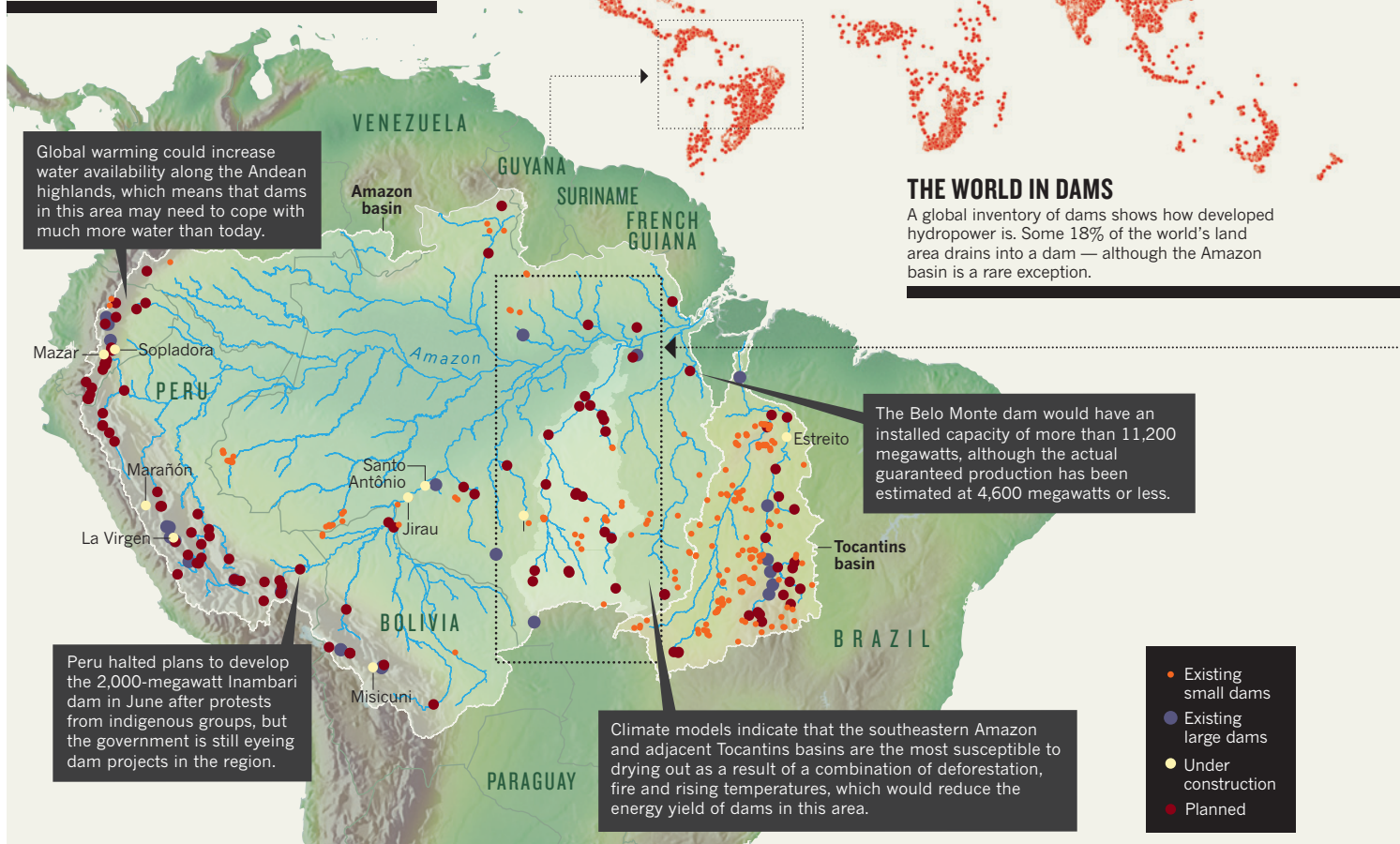
ON THE BLOG

Bat flight evolved from a flutter to get at insects, rather a glide
go.nature.com/xgodno



A BECKONING PRIZE

With thousands of kilometres of free-flowing waterways, the Amazon region looks increasingly attractive as an energy source for Brazil and other nations.



ENVIRONMENT

A struggle for power

Brazil is developing the last great untapped reserve of hydroelectricity, the Amazon basin.

BY JEFF TOLLEFSON

When a few hundred demonstrators, mostly from indigenous communities, temporarily occupied the construction site of the Belo Monte dam on Brazil's Xingu River early on 27 October, workers laid down their tools. But the Brazilian government did not back down from its stance that this hydroelectric project on a tributary of the Amazon — expected to be among the world's largest, with a capacity of 11,000 megawatts, when completed in 2015 — is essential to meeting the energy needs of a booming economy. Under a court order, the demonstrators vacated the site later the same day, but the dam remains the subject of fierce litigation.

The episode briefly drew the world's

attention to a controversial mega-project, but this is only part of a larger picture. Led by Brazil, governments in the region are increasingly looking to tap into the Amazon system to slake a growing thirst for energy. If current plans are realized, a wave of dam construction will bring staggering change and development to the rainforest in the coming decades.

In a global context the Amazon stands out as an area of untapped potential, with the world's greatest river system and a paucity of hydroelectric stations, says Mark Mulligan, a geographer at King's College London, who has led the development of an interactive database of more than 36,000 dams around the world. One of his former students, Leonardo Sáenz, has moved on to Conservation International in Arlington, Virginia, where he is improving the database and

incorporating dams that are planned and under construction in the Amazon (see 'A beckoning prize'). The goal is to understand how those investments affect the broader landscape, both physically and economically.

According to the conservation group WWF, less than 10% of Brazil's electrical power comes from dams in the Amazon region at present. The Belo Monte dam would boost this figure, and many more projects are on the drawing board, including 18 dams proposed for the Tapajós tributary system alone over the coming decade. Brazil has also signed an agreement to develop hydroelectric dams in the Peruvian Amazon in exchange for a share of the power.

Although the dams promise carbon-free electricity, they also lead to more road construction and deforestation as well as invasions

LITIGATION

US lawsuit extends thalidomide's reach

Drug blamed for a broader range of harmful effects.

BY MEREDITH WADMAN

In a new twist of a historic tragedy, 13 Americans who say they are survivors of thalidomide are suing four companies for producing and distributing the notorious drug. They say that the drug — used by pregnant women for morning sickness until it was discovered to cause severe birth defects — affected more people in the United States than thought, and caused a wider range of deformities. And, they say, the companies have done all they can to hide these facts.

Thalidomide's devastating effects first came to light 50 years ago this month in the German newspaper *Welt am Sonntag*. In Europe, the drug was implicated in thousands of cases of malformed newborns, but in the United States the damage was limited because the Food and Drug Administration (FDA) refused to approve it for market. Until now, most US cases were thought to originate from thalidomide obtained abroad.

The lawsuit, filed in a Philadelphia court on 25 October, asserts that before thalidomide was pulled from markets around the world, samples were doled out to more than 1,200 physicians in the United States by three companies whose legal liabilities are now the property of Sanofi-Aventis US, based in Bridgewater, New Jersey. Separately, it alleges, citing an FDA memorandum that only came to light earlier this year, the company Smith, Kline & French, now GlaxoSmithKline (GSK), ran a clinical trial of the drug in the US involving 875 people, including pregnant women, in 1956–57. The suit claims that at least one malformed baby was born to a trial participant in 1958. (The German firm Grünenthal, based in Aachen, and Avantor Performance Materials, based in Center Valley, Pennsylvania, are also named in the suit.)

Sanofi-Aventis and Grünenthal say that they cannot comment on ongoing litigation.

Mary Anne Rhyne, a spokeswoman for GSK, says that the company “intends to vigorously defend itself against this lawsuit”. She notes that Smith Kline & French never manufactured or sold thalidomide and adds: “The Plaintiffs’ complaint is replete with scientific inaccuracies and misstatements.”

The company challenges the claim that thalidomide can cause limb defects that are confined to one side of the body, as seen in

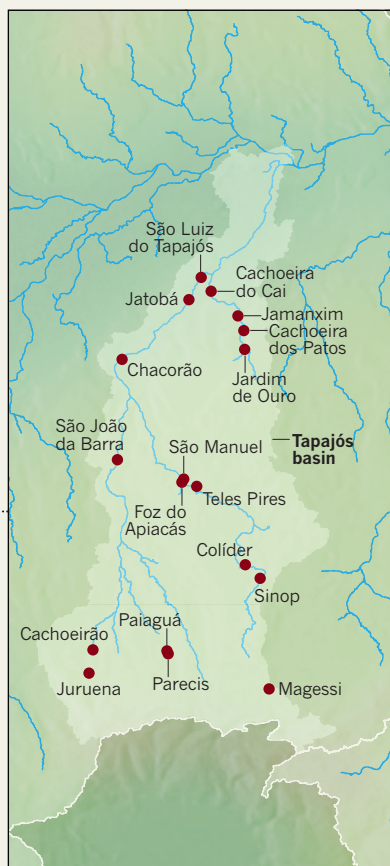
nine of the plaintiffs. Conventional wisdom has long held that thalidomide’s signature defect — a shortened, seal-like ‘flapper’ arm, known as phocomelia — affects both sides of the body.

The plaintiffs’ lawyers argue that this assumption is unproven. “There are no representative, controlled studies documenting the true spectrum of thalidomide injuries,” they write in the lawsuit. “A universe of thalidomide related injuries has been thereby excluded from diagnosis.” They further suggest that “recently available studies published in medical and scientific journals reveal the flaws in the orthodox medical opinion”.

When asked by *Nature* for relevant studies, the plaintiffs’ lawyers at Hagens Berman Sobol Shapiro in Seattle, Washington, pointed to work showing one-sided limb defects in chick embryos exposed to thalidomide and thalidomide analogues (C. Therapontos *et al.* *Proc. Natl Acad. Sci. USA* **106**, 8573–8578; 2009). The paper’s senior author, Neil Vargesson of the School of Medical Sciences at the University of Aberdeen, UK, says that the one-sidedness was due to the physical orientation of the developing chick when the medication was injected into the egg.

Vargesson says his work does not confirm or deny that the plaintiffs’ defects are the result of thalidomide. “The biggest issue facing the lawyers is persuading authorities that thalidomide gave rise to a range of other defects, including unilateral limb defects — or that it caused other damage without apparent limb defects at all.” However, adds Vargesson, who has advised lawyers for potential plaintiffs in the United Kingdom who do not have apparent limb defects, “it’s pretty clear that this drug did an awful lot of things and they don’t always centre around limb defects”.

Lewis Holmes, an expert in birth defects at Massachusetts General Hospital in Boston, says that the plaintiffs will have an uphill struggle to support their argument from the scientific literature because of the lack of systematic studies that follow up the offspring of women who took thalidomide during pregnancy. Holmes also notes that the relative paucity of thalidomide births in the United States means that few researchers there can speak with authority on the drug’s effects. “None of us ever saw thalidomide-damaged children,” he says. ■



TAPAJÓS BASIN

Among the projects most likely to move forward are several located within the Tapajós river basin, where the Brazilian government has focused its planning efforts over the next decade.

of migrant workers and massive methane emissions when large swathes of forest are drowned. And, increasingly, experts fear that changing patterns of rainfall brought about by deforestation and climate change could reduce the energy return from dams, rendering many investments obsolete.

“It’s really easy to get your infrastructure wrong, and that poses serious investment risks in the long run,” says John Matthews, a freshwater expert at Conservation International. “From this perspective climate change presents the ultimate risk in the Amazon.” Matthews fears that Brazil could become perilously reliant on an uncertain energy source, even as the government builds more dams.

“They are opening a new hydropower frontier, the last hydropower frontier in South America,” says Pedro Bara, who works for the WWF in Brasília. “In 30 years, if all of the plans were implemented, half of Brazil’s energy would come from the Amazon.” ■

➔ **NATURE.COM**

Read about dams in southeast Asia: go.nature.com/ymuzel

GLOBAL HEALTH

Aid organizations tap into social-science expertise

Behavioural and cultural studies seen as key to success of public-health initiatives.

BY ERIKA CHECK HAYDEN

Condoms. Breastfeeding. Hand washing. After a decade of ramped-up spending, the donors to global-health programmes are grappling with a thorny issue: why are some health-care interventions that are proven life-savers failing to save as many lives as they should in the field, even after unprecedented investment to support them?

The question topped the agenda at a meeting last week of industry executives, researchers, public-health workers and global-health campaigners convened by the Bill & Melinda Gates Foundation in Seattle, Washington. The foundation is looking to enlist insights from the behavioural and social sciences to help in two key areas: helping communities to adopt proven interventions, and scaling these measures up to help more people in poor countries.

In the past, the foundation has invested heavily in developing technological innovations but comparatively little on persuading people to embrace them. It is now changing that by commissioning reports on how to build social and behavioural research into future development programmes to make them more successful. Foundation officials say that they are also considering these issues at each step in programmes already under way. The changes were spurred by discouraging reports such as one published in September (R. Lozano *et al. Lancet* **378**, 1139–1165; 2011). This found that only 9 out of 137 developing countries are on track to meet two ambitious targets under the United Nations' Millennium Development Goals: to reduce maternal deaths by three-quarters and child deaths by two-thirds by 2015.

Many treatments, such as oral-rehydration therapy to treat diarrhoea, or 'kangaroo mother' care, in which premature infants are held in skin-to-skin contact with their carer, are inexpensive and simple, but cultural and behavioural factors have slowed their adoption. "There isn't anything we do in global health that doesn't have some component of social change," Melinda

"We may have the best malarial bed net, but if people don't sleep under it, it's not going to make a difference."



Oral rehydration is a cheap and simple life-saver.

Gates told delegates in her opening address. "We may have the best malarial bed net, but if people don't sleep under it, it's not going to make a difference."

Participants at the meeting cited numerous examples of how common sense and local knowledge could help to overcome barriers. Rashad Massoud, director of the USAID Health Care Improvement Project in Bethesda, Maryland, described a programme in Niger that aimed to prevent deaths from post-partum haemorrhage using the drug oxytocin. The programme was failing because oxytocin, which must be refrigerated, was delivered at night, when health-centre pharmacies were closed and could not accept deliveries. Public-health workers urged local communities to come up with their own solutions, such as using coolers to store the medicine, and maternal deaths from haemorrhage plummeted.

"A lot of the solutions we've figured out have come from communities themselves," noted Nana Twum-Danso, who directs African operations for the non-profit Institute

for Healthcare Improvement in Cambridge, Massachusetts. She told the story of a project aimed at convincing women in Ghanaian villages to go to clinics to have their babies. Project leaders learned that women were reluctant to do this because it was perceived as cowardly. The villagers themselves came up with solutions to the problem, one of which was to fine men whose wives gave birth at home.

Gary Darmstadt, director of family health at the Gates Foundation, said that marketing expertise could also help. The private sector has experience in assessing markets' openness to new products and finding ways to get consumers to demand those products, he says. Such an approach might help the foundation to achieve its goals in cases such as promoting use of the antiseptic chlorhexidine to cleanse a newborn's umbilical cord to prevent infection. This has proved a challenge because many cultures have traditions surrounding the umbilical cord, such as smearing it with cow dung, oils or ash. "We're looking at how we design a product that makes people want to buy it and put it on, and these are the kinds of things that public health really hasn't done," Darmstadt said.

At the meeting, Donald Berwick, head of the US Centers for Medicare and Medicaid Services, in Baltimore, Maryland, praised the foundation for focusing attention on how to scale up proven health-care interventions. This could inspire a worldwide change, even in developed countries such as the United States, whose own health-care system is in "crisis", he said.

"If the Gates Foundation is interested in this, it will change the world," Berwick said. ■

CORRECTIONS

The News story 'Mental-health guide accused of overreach' (*Nature* **479**, 14; 2011) wrongly stated that the DSM-5 petition was sponsored by the American Psychological Association. It was co-sponsored by five divisions of the association, not the association as a whole.

The News Feature 'The Black Death decoded' (*Nature* **478**, 444–446; 2011) wrongly located Sharon DeWitte at the State University of New York. She is now at the University of South Carolina in Columbia.

G. PIROZZI/PANOS

The quake killer

The US government says that a huge earthquake risk lurks in the heart of the country, where a series of large shocks hit 200 years ago. Seth Stein says that kind of warning is dead wrong.

BY RICHARD MONASTERSKY

The lethal fault cuts through the middle of a Tennessee bean field and then ducks beneath the Mississippi River, making a beeline for New Madrid, Missouri. Named the Reelfoot fault, this geological crack combined with neighbouring faults two centuries ago to unleash a series of devastating earthquakes that have been called the biggest to strike the contiguous United States in recorded history. On government hazard maps, the New Madrid region stands out as a red bull's eye. This spot in the middle of the continent — far from the plate boundaries that produce Earth's greatest quakes — would seem to be every bit as dangerous as San Francisco or Los Angeles.

Seth Stein poses for a picture on top of the Reelfoot fault, but the geology just doesn't cooperate. For something supposedly so dangerous, the quake-maker is hardly noticeable: the only sign is a 3-metre-high bump. "You don't see much of a fault," snorts Stein, a geophysicist at Northwestern University in Evanston, Illinois, who has studied the region for the past 20 years.

To Stein, the lack of anything substantial to photograph is one of several pieces of evidence suggesting that the US government and many scientists are wrong about the hazard here. If the Reelfoot fault had been popping off giant earthquakes repeatedly, for a long geological time, it would have built up a noticeable scarp, like the impressive topography of a mountain range. But there's nothing like that here. And when Stein and his team surveyed the region with Global Positioning System (GPS) instruments, they found no evidence that the ground was warping in preparation for a repeat of the New Madrid quakes.

The seismic-hazard maps that show the New Madrid region to be so dangerous, he says, are sort of "an emperor's-new-clothes business. Our role was to point out the emperor has no clothes." The consequence, says Stein,

➔ NATURE.COM
For a podcast and
slideshow on this
story, see:
go.nature.com/xz9con

is that the government and businesses in cities such as Memphis, Tennessee, are mispending vast sums — potentially

billions of dollars — to construct buildings strong enough to withstand a giant quake that will never show up. "Building the Midwest to California construction codes is a colossal waste of money," he says.

Stein and his group have published papers in *Nature*, *Science* and other journals presenting evidence to support that idea, and last year Stein took his case to the public with a book called *Disaster Deferred*, which accused the US Geological Survey (USGS) of inflating the quake risk in the region, known as the New Madrid Seismic Zone. That charge, coupled with Stein's habit of ridiculing people and ideas he finds absurd, has irritated many geoscientists. But his work has forced them to confront the question of whether the United States is over-preparing for earthquakes in New Madrid, while under-preparing elsewhere. The magnitude-5.6 quake in Oklahoma this week and a 5.8 in Virginia in August point to the potential for strong shocks throughout central and eastern United States.

As he tours the New Madrid region with

Seth Stein argues that money is being wasted preparing the American Midwest for a major earthquake that may never come.

Over the next few weeks, hundreds of aftershocks rattled the region, followed by two more major shocks on 23 January and 7 February 1812. The February quake caused the land southwest of the Reelfoot fault to jerk up, whereas New Madrid and the rest of the land northeast of the fault dropped by a combined total of about 3 metres or more. The quake destroyed the town and the Mississippi soon swallowed the land, forcing residents to rebuild farther from the river. The vibrations from the quakes were felt across an area of more than 2.6 million square kilometres — about one-third of the area of the contiguous United States (see ‘Earthquake central’).

Because the quakes happened before the advent of modern seismometers — and in such a sparsely settled part of the young country — researchers have struggled over the years to determine just how big they were. In 1996, Arch Johnston, a geophysicist at the University of Memphis, used historical accounts of the quakes to estimate that the first shock, which he thought was the largest, had a magnitude somewhere between 7.8 and 8.4 — a truly monstrous earthquake, bigger than any in the history of the United States outside of Alaska¹.

The events sound even worse in the museum, where a film calls the largest shock “the worst earthquake in recorded history”. Virtually the entire North American continent experienced some shaking, claims the film.

A museum pamphlet cites a study, sponsored by the Federal Emergency Management Agency, projecting that 3,500 people would die if a magnitude-7.7 earthquake were to strike the New Madrid region today. The economic losses, according to the pamphlet, could exceed US\$300 billion. Stein grimaces at the film and scoffs at what he considers hazard-mongering. “The scary thing is how much of this is completely wrong,” he says. “The hazard here is wildly overestimated.”

FINDING FAULT

It is that kind of pronouncement that wins Stein few points with his rivals. In debates, his attacks on positions he feels are unsupportable can be hyperbolic. And when his rants reach full steam, he seems like a hyper-intelligent version of George Costanza, the ever-complaining character from the television show *Seinfeld*, who, like Stein, also happens to be short and bald with glasses.

But Stein has a fertile imagination and has made several important advances in geophysics. As a young faculty member at Northwestern in the 1980s, he and his colleagues developed a widely used model quantifying how Earth’s plates have moved over the past 3 million years. The model

JEAN-MARC GIBOUX/GA BY GETTY IMAGES

Nature, Stein lays out his case. The evidence emerging from New Madrid, he says, suggests that the faults involved in the quakes of 1811–12 are shutting down, shifting the hazard to other faults perhaps hundreds of kilometres away. Stein has hypothesized that the centres of other continents behave in a similar way, with patches of earthquake activity that harass one locale for thousands of years and then jump to other regions.

Some of his arguments are gaining ground, although few researchers go as far as Stein in playing down the dangers in the New Madrid area. “What he’s doing here is good; he’s raising issues and making hypotheses that could well be true,” says John Vidale, a seismologist at the University of Washington in Seattle, who this year led a federal panel that investigated the hazard in New Madrid. “The only thing that angers the community is when he argues that he’s right and everybody else is wrong.”

New Madrid is a quiet outpost on the banks of the Mississippi, where the mighty river meanders briefly north and then makes

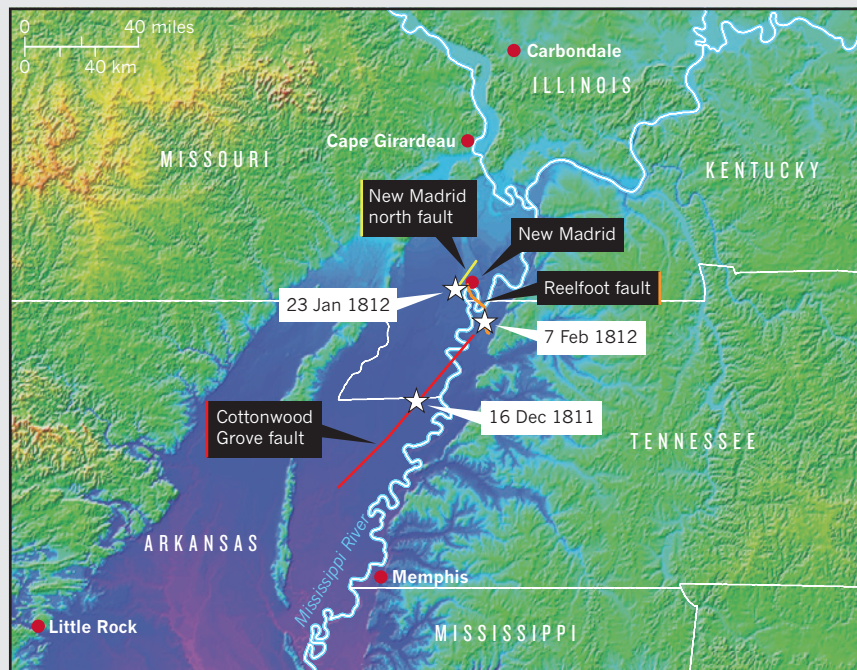
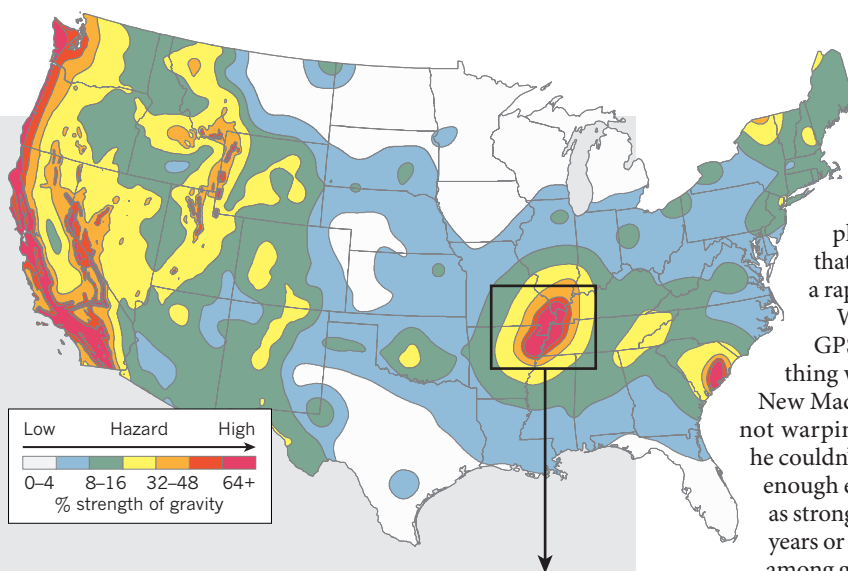
a complete U-turn back south. Throughout much of its history, the town fought — and lost — epic battles with the ever-shifting river, until levees were built in the early twentieth century.

The New Madrid Historical Museum sits next to the levee and honours the town’s seismic history with a detailed exhibition, currently being upgraded in advance of the quakes’ bicentennial next month. At the time of the shocks, New Madrid was a bustling port — one of the most important on the lower half of the Mississippi River.

The devastating seismic sequence started with a thunderous bang at around 2 a.m. on 16 December 1811. According to one resident: “We were awakened by a most tremendous noise, while the house danced about and seemed as if it would fall on our heads.” The quake and its immediate aftershocks scared many of the residents deeply, but few were killed. “Several men, I am informed, on the night of the first shock deserted their families, and have not been heard of since,” wrote the New Madrid resident.

Earthquake central

Three major earthquakes struck near the town of New Madrid, Missouri, during the winter of 1811–12. From historical accounts and studies of the geological scars left by the quakes, researchers hypothesize that they occurred on three separate faults (below). Because of the region's history, it stands out on a map of seismic hazard prepared by the US Geological Survey.



motion is slower and slower,” admits Robert Smalley Jr, a geophysicist at the University of Memphis, who reported in 2005 that the region was straining at a rapid rate⁴.

When Stein saw Newman's GPS analysis, he knew something was wrong with the whole New Madrid story. If the crust was not warping in any noticeable way, he couldn't see how it could store up enough energy to fuel earthquakes as strong as magnitude 8 every 500 years or so — the accepted wisdom among geoscientists at the time. So

Stein and his team proposed that the quakes of 1811–12 were much smaller than previously estimated, perhaps no larger than magnitude 7, which, because the scale is logarithmic, would produce just one twenty-fifth the shaking of a magnitude-8.4 quake.

As with the GPS results, other research has since backed up Stein's estimate. When Susan Hough, a seismologist at the USGS in Pasadena, and her colleagues went back to witnesses' reports of the 1811–12 quakes, they estimated that the biggest shock was between magnitude 7.4 and 7.5 (ref. 5). This year, she downgraded the size even further⁶, suggesting that somewhere between 6.8 and 7.0 would fit the data best, which is around what Stein had suggested more than a decade earlier.

Yet even before all the confirming evidence emerged, Stein had pushed his GPS data to a controversial conclusion. The New Madrid zone might well be turning off, after a geologically brief spell of activity, he claimed — and large quakes like the 1811–12 sequence might never happen there again.

assumes that each plate is a rigid patch of Earth's outer shell, with not much movement in its interior. The vast majority of earthquakes, after all, happen at the edges of plates, where they crunch together and warp into mountain ranges. But Stein wondered how much the plates were deforming in their centres.

The US system of GPS satellites gave him a means of finding out. In 1991, he began a study of the New Madrid area to track the warping of the crust in the centre of the North American plate.

Stein was certain that the crust was deforming to some degree there because geologists had been finding evidence of a series of large quakes hundreds of years before the cluster in 1811–12. The energy for those quakes must have come from forces straining the crust, perhaps transmitted all the way from the edges of the North American plate. But before Stein could get results from his GPS study, a rival group beat him

to the finish by taking a shortcut. The team compared one set of GPS measurements with land-surveying data from the 1950s and found that the New Madrid area was deforming at a rapid rate, about one-third as fast as the San Andreas fault in California².

Because they were relying on GPS data alone to catch any movement, Stein and his team had to wait several years and then resurvey the suite of sites around New Madrid, which they did in 1993 and 1997. When Andrew Newman, a graduate student working with Stein, finally crunched the numbers, he found something odd. There was little — or possibly no — warping of the crust³. “I was utterly flabbergasted that we did not see anything,” says Stein.

For several years, research teams battled in the scientific literature about whether the crust was deforming around New Madrid. But as more data piled up, they corroborated Stein's original finding. “As we get better and better at measuring, we're finding this

UNDER PRESSURE

Like most faults across the stable part of North America, the ones in the New Madrid area are hundreds of millions of years old and have remained quiet for most of their lives. But something must have activated them in the recent geological past — within the past 20,000 years — and geoscientists have been struggling to understand what that was.

A potential answer can be found in the tawny bluffs along Pawpaw Creek in northwest Tennessee. On a visit there, Stein and geologist Roy Van Arsedale of the University of Memphis examine a 25-metre cliff that provides a superb view of the sedimentary strata that are normally hidden beneath the surface. Farther west of this site, in the Mississippi River Valley, parts of this layer-cake sequence are missing. The river stripped away 12 metres of sediment at the end of the last ice age, between 16,000 and 10,000 years ago.

In a study of the sedimentary sequence throughout the area, Van Arsedale and his

colleagues found that the erosion of those sediments happened right above the faults that have been implicated in the New Madrid earthquakes. They proposed that the removal of all that weight altered the stresses in the subsurface rocks enough to reactivate the New Madrid faults⁷.

To test that idea, Van Arsdale and Stein worked with Eric Calais and Andrew Freed at Purdue University in West Lafayette, Indiana, to model the effects of removing so much weight from above the faults. They calculated that the erosion unclamped the faults enough for the ones on the verge of an earthquake to fail, releasing energy that had accumulated over the past — perhaps millions of years earlier during events such as the growth of mountains in western North America⁸. But once a particular fault has broken, the scientists suggested, there would not be enough stress in the region to trigger another big quake there. So the large prehistoric quakes in that region must have happened on different fault segments from the ones that broke two centuries ago.

Rather than seeing the faults around New Madrid in isolation, Stein suggests that they are part of a broad system of interacting faults, and that seismic activity shifts from one set of faults to others over hundreds or thousands of years. He sees a similar pattern in Europe, Australia and China, where earthquakes often do not recur in the same spots.

If the New Madrid area were to shut down, the seismic activity would eventually transfer somewhere else, says Stein, probably to the southwest or to the Wabash Valley fault system in Indiana and Illinois, which had a major earthquake 6,000 years ago. Van Arsdale is concerned about the possibility of quakes shifting south towards Memphis, along the edge of the Reelfoot rift, where the North American continent began, briefly, to pull apart some 500 million years ago.

Stein, Van Arsdale and others would now like to get GPS measurements and construct a detailed earthquake history of a much broader region around New Madrid, to determine how seismic patterns have moved

“How on Earth did the survey make this the most dangerous place in North America?”

among the different faults in the central United States. “Compared with ten years ago, we’ve made a hell of a lot of progress,” says Stein. “I think we’re now poised to make significant advances.”

The evidence already available, however, is enough to undermine the government’s assessment of the earthquake danger in the New Madrid area, Stein says. “How on Earth

RISKY SCIENCE

Do maps magnify the earthquake hazard?

On official seismic-hazard maps, the midsection of the United States looks as risky as the quake-prone West Coast (see ‘Earthquake central’). That’s because the Missouri area was hit by major quakes in 1811–12 and because there is concern about another large quake. But some researchers argue that the assumptions that go into the hazard maps — and the way that the maps are produced — make the region look even more dangerous.

Used to set building requirements, the hazard maps created by the US Geological Survey (USGS) represent estimates of the probability that ground shaking in a given area will exceed certain levels.

Arthur Frankel, a seismologist with the USGS in Seattle, Washington, who led a 2002 assessment, says that the New Madrid area has a high hazard largely because seismic waves travel efficiently through the hard rock of the central and eastern United States. So the rare, large earthquakes in New Madrid affect a much larger area than do similarly sized quakes in California.

But the criteria used to make the maps also have a role. The USGS produces several types of maps that estimate the strength of shaking in a given area from all potential seismic sources. The agency — like the engineers who set building codes — favours maps that show the levels of shaking that have a 2% chance of being exceeded in a 50-year period. These correspond to relatively rare but large events. Another set of

maps estimates the strength of shaking that has a 10% chance of being exceeded in 50 years, which reflects the kind of earthquakes that strike more frequently. The New Madrid area looks less hazardous in the 10% maps than in the 2% maps, because the geological evidence suggests that clusters of large quakes in that region might occur at intervals of 500 years or more, whereas smaller ones happen more frequently.

Seth Stein, a geophysicist at Northwestern University in Evanston, Illinois, has argued that the survey — and building codes — should work from the maps that depict levels that have a 10% chance of being exceeded in 50 years; these are the hazard probabilities used in Europe to set building codes and they better represent the risks that people and structures will probably face in the near future, he says.

But that can lead to problems, says Gail Atkinson, a seismologist at the University of Western Ontario in London, Canada. Atkinson helps develop building codes for Canada, which uses the 2%-in-50-year estimates. She notes that New Zealand designs for the 10%-in-50-year shaking levels and suffered tremendous losses from an earthquake in Christchurch in February, which killed 181 people. “I don’t think most people in New Madrid would be pleased with that sort of an outcome,” she says. If the country had prepared for extreme, rare events, then more buildings might have withstood the shock. **R.M.**

did the survey make this the most dangerous place in North America?” he asks. “It’s a relatively small hazard.”

This year, the National Earthquake Prediction Evaluation Council, which advises the USGS, convened a panel of independent

“it is likely that the estimated New Madrid Seismic Zone hazard may decline moderately in the next hazard assessment due to improved knowledge of past earthquakes and current deformation” (see ‘Do maps magnify the earthquake hazard?’).

According to Vidale, the main problem is that the region remains poorly understood. “We basically found that there’s a lot of uncertainty. Because we don’t know exactly what’s going on, we need to be prepared for large earthquakes.” But he also acknowledges the possibility that a large earthquake might never come.

About Stein, Vidale says: “None of his ideas are bad. It’s just that he presents them in an extreme way and doesn’t allow for the possibility that he might not be right.”

Stein’s overall argument gets little backing even from Hough, who has argued strongly that the earthquakes 200 years ago were much smaller than once thought. “Seth argues that the zone is dead because you don’t see strain.

scientists and engineers to examine the issue. Vidale, who led the panel, said they were motivated in part by Stein’s persistent criticism as well as by the impending bicentennial of the quakes. The panel in general supported the survey’s most recent assessment from 2008, which showed that the hazard in parts of the New Madrid region surpasses that in Los Angeles or San Francisco, although it said



USGS

Pictures taken a century after the 1811–12 quakes show their lasting effects. The shaking liquefied buried sand layers, causing them to erupt through fissures (left) and cover the ground in thick deposits that smothered trees (right).

And I don't think that's a valid conclusion," she says. "There's enough strain around that could produce a magnitude 7."

Stein, though, denies that he has been so dogmatic. He says that, in his book and in papers, he presents the idea of New Madrid turning off as a possibility — not a certainty.

REALITY CHECK

These issues all come to a head in downtown Memphis, which is full of office buildings, apartments and hotels constructed during the first third of the twentieth century. Many of them are unreinforced masonry — made of concrete blocks or bricks without any metal skeleton. Engineers consider this class of structure to be one of the most vulnerable during an earthquake. If a quake near magnitude 8

Tomasello, an engineer with the Reaves Firm in Memphis who has collaborated with Stein. But many people in the area have balked at such stringent requirements, and the city is allowing owners to build to a less restrictive code until it adopts the 2009 version of the IBC, which is expected to happen around the end of this year.

Some engineers suggest that the code's enhanced seismic provisions would increase building costs only marginally, but Tomasello contends that they could add an extra 10–15%. If all new construction were required to meet the newer code, building costs in the Memphis metropolitan area would be increased by \$200 million a year, he estimates. "When you look at the hazard, it just doesn't make any sense," says Tomasello.

government to do a rigorous analysis of the costs and benefits of seismic building codes, before forcing people to foot the bill for expensive protections.

The next day, as Stein walks through Memphis airport on his way back home, he brings up the issue again. It is 13 September, just two days after the tenth anniversary of the 2001 terrorist attacks on the United States. In the intervening decade, the country has spent \$1.3 trillion waging two wars and has instituted time-consuming security provisions at airports, such as herding travellers through scanning machines capable of peering through clothes.

To Stein, these are all examples of misunderstood risks and squandered resources. "I can't do anything about the thousand things the government does that I'm unhappy about," he says in a quiet moment. But with New Madrid, "I can do something about that." ■

"If you spend that money on schools or health care, you're almost guaranteed to save lives."

were to strike within the next few decades, it could topple many of these buildings. But if no large quake comes for 50 years or more, most of these buildings will probably be long gone, replaced by newer structures.

The Federal Emergency Management Agency has been pushing municipalities in the Memphis area to adopt a standard International Building Code (IBC) that would require new buildings to withstand 3 to 4 times stronger shaking than was required by earlier codes, says Joseph

To Stein, building to a lower level of protection makes sense. The faults that produced the 1811–12 quakes are still crackling with aftershocks, and there is a chance that one of them could be as large as magnitude 6.7. But building to the new IBC in Memphis would be a waste of money, he says. "There's a good chance that all the money you would spend would save no lives at all. But if you spend that money on schools or health care, you're almost guaranteed to save lives." The most important thing, Stein says, is for the

Richard Monastersky is a features editor for *Nature* in Washington DC.

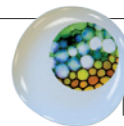
1. Johnston, A. C. *Geophys. J. Int.* **126**, 314–344 (1996).
2. Liu, L., Zoback, M. D. & Segall, P. *Science* **257**, 1666–1669 (1992).
3. Newman, A. *et al. Science* **284**, 619–621 (1999).
4. Smalley, R. Jr, Ellis, M. A., Paul, J. & Van Arsdale, R. B. *Nature* **435**, 1088–1090 (2005).
5. Hough, S. E., Armbruster, J. G., Seeber, L. & Hough, J. F. *J. Geophys. Res.* **105**, 23839–23864 (2000).
6. Hough, S. E. & Page, M. J. *J. Geophys. Res.* **116**, B03311 (2011).
7. Van Arsdale, R., Bresnahan, R., McCallister, N. & Waldron, B. *Geol. Soc. Am. Spec. Pap.* **425**, 177–192 (2007).
8. Calais, E., Freed, A. M., Van Arsdale, R. & Stein, S. *Nature* **466**, 608–611 (2010).

COMMENT

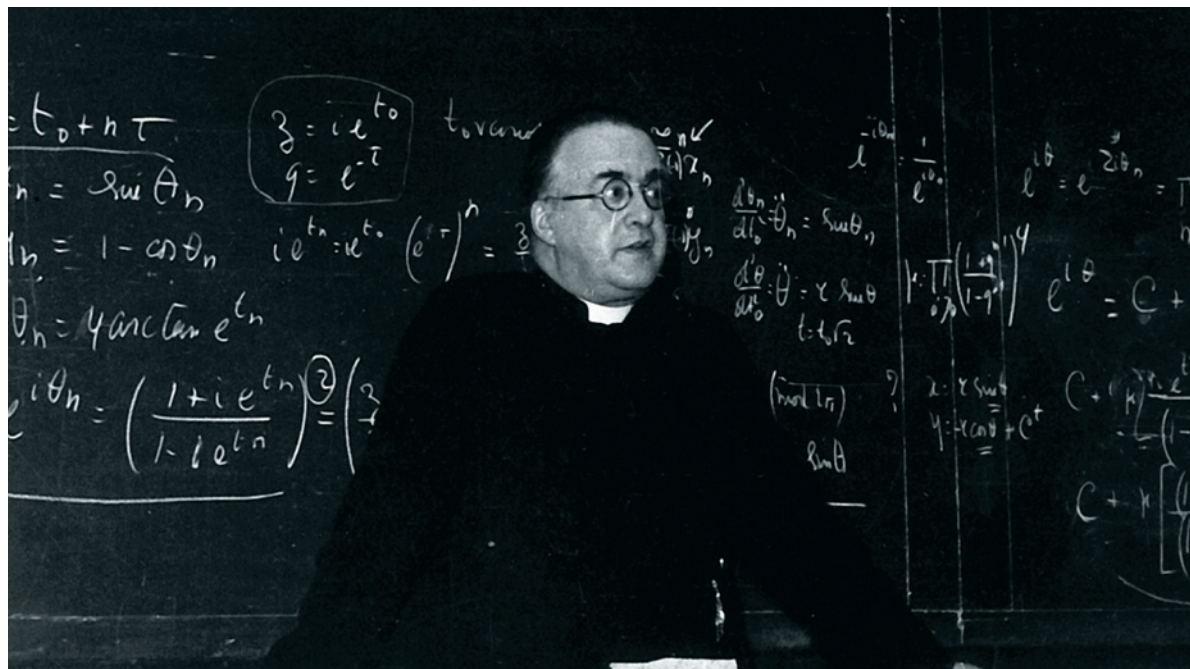
ART Scientific clues verify a rediscovered painting by Leonardo da Vinci **p.174**

PSYCHOLOGY How to persuade and influence people **p.176**

BOTANY Microscopic views of plants grace ceramics **p.177**



NOBELS Immunologists question priority for this year's prize in medicine **p.178**



Georges Lemaître giving a lecture at the Catholic University of Louvain in Belgium.

Mystery of the missing text solved

A discovered letter explains the loss of key paragraphs during the translation of one of Georges Lemaître's papers about the expanding Universe, shows **Mario Livio**.

A passionate debate has flared up in recent months about who deserves the credit for one of the most profound discoveries of our time: that our Universe is expanding, and so had a beginning¹⁻³. The American astronomer Edwin Hubble, who tracked the expansion in the velocities and distances of scores of distant galaxies during the 1920s, is usually cited. But a few articles have raised the suspicion that someone censored a key paper by the Belgian priest and cosmologist Georges Lemaître to ensure Hubble's priority^{2,3}.

There is little doubt that Lemaître deserves the credit for proposing an expanding Universe. But the censorship charges tarnish Hubble's genuine achievement of confirming and extending the idea. As someone intimately involved with Hubble's namesake — the Hubble Space Telescope — I became intrigued by this 'whodunnit' mystery, and decided to investigate. As a result, I unearthed

NATURE.COM
Edwin Hubble in translation trouble:
go.nature.com/hca654

a letter from Lemaître that, to my satisfaction, ends the debate.

Here are the background facts. By February 1922, American astronomer Vesto Slipher had measured the redshifts (frequency shifts indicating relative motions) for 41 galaxies (then known as nebulae) in the northern sky. Listing them in his 1923 book *The Mathematical Theory of Relativity*, British physicist Arthur Eddington noted that: "The great preponderance of positive [receding] velocities is very striking." But he added that a lack of observations of southern nebulae prevented him from drawing further conclusions.

In 1927, Lemaître published, in French, a remarkable paper in the relatively obscure *Annals of the Scientific Society of Brussels*¹. It was entitled (in its English translation): 'A homogeneous Universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae'. In it, Lemaître reported his discovery of dynamic solutions to Einstein's general relativity equations, from which he derived what

ARCHIVES GEORGES LEMAÎTRE/CATHOLIC UNIV. LOUVAIN/TECLIM

► is now known as Hubble's law — that the velocity at which a galaxy appears to recede is proportional to its distance from us.

But Lemaître went beyond theoretical calculations in the paper. He determined the rate of expansion of the Universe using the velocities of the galaxies measured by Slipher (and published⁵ by Gustaf Strömberg, a Swedish astronomer at the Mount Wilson Observatory in California), and the distances to them as determined from brightness measurements published by Hubble⁶ in 1926. For the value of that rate, today called the Hubble constant, Lemaître obtained 625 kilometres per second per megaparsec. Lemaître also commented in the paper that the accuracy of the distance estimates available at the time was insufficient to assess the validity of the linear relation he had discovered.

Two years after Lemaître's paper appeared, Hubble published a paper⁷ entitled 'A relation between distance and radial velocity among extra-galactic nebulae'. In it, he and his assistant, Milton Humason, used improved distances (in part based on better stellar distance indicators, such as Cepheid variables and novae) and velocities taken mainly from Slipher, to establish the existence of Hubble's law, and to determine a value for the Hubble constant of 500 kilometres per second per megaparsec.

On the basis of this story, it would seem fair to credit the discovery of the expanding Universe and the tentative existence of a Hubble law to Lemaître; and the detailed confirmation of that law to Hubble and Humason, given their subsequent meticulous observations, which extended Slipher's velocity measurements to greater distances. But here the plot thickens.

LOST IN TRANSLATION

The English translation of Lemaître's 1927 paper was published in the *Monthly Notices of the Royal Astronomical Society* in March 1931 (ref. 8). However, during the process, a few paragraphs from the original French version were deleted, notably the one in which Lemaître described Hubble's law and derived the expansion rate.

Also missing were a paragraph in which Lemaître discussed errors in the distance estimates, and footnotes, in one of which he interpreted the proportionality between the velocity and distance as resulting from a cosmic expansion. In the same footnote, Lemaître calculated two possible values for the Hubble constant, of 575 and 670, depending on how the data were grouped.

That these paragraphs are missing from the translated paper has been known for some time, although not widely. Cosmologist Jim Peebles at Princeton University in New Jersey, noted in a volume on Lemaître in 1984 that: "It is curious that the crucial paragraphs describing how

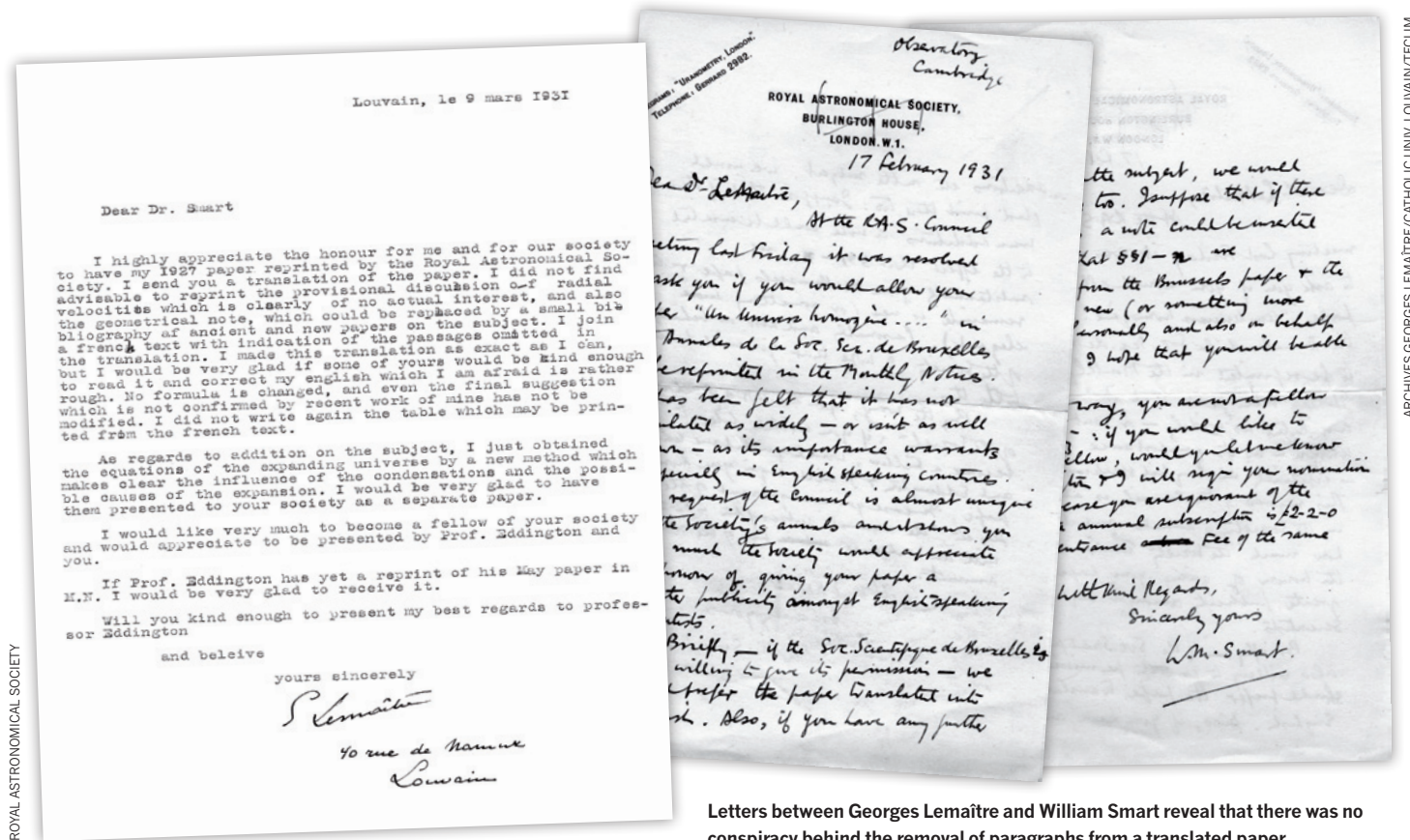
Lemaître estimated H [the Hubble constant] and assessed the evidence for linearity were dropped from the 1931 English translation."

Who translated Lemaître's paper and why were these paragraphs deleted? Canadian astronomer Sidney van den Bergh speculated earlier this year that whoever did the 'selective' editing may have done so to prevent Lemaître's paper from undermining Hubble's priority claim⁹. David Block, a

mathematician at the University of the Witwatersrand in Johannesburg, South Africa, suggested³ further that Hubble might have had a hand in this cosmic censorship, to ensure that credit would go to himself and to the Mount Wilson Observatory, where he made the observations. Historian of science Robert Smith at the University of Alberta in Edmonton, Canada, who also believes that most of the credit for discovering the expanding Universe should go to Lemaître, has suggested that the paragraphs may have been removed as part of standard editorial practice by the editor of the *Monthly Notices*.

Wanting to find out more, I examined original documents linked to the paper. With the help of Liliane Moens from the Archives Georges Lemaître in Louvain, Belgium, I obtained a copy of the letter sent by the then editor of the *Monthly Notices*, astronomer William Marshall Smart, to Georges Lemaître, concerning the translation and publication

"This clearly ends speculation about who translated the paper and who deleted the paragraphs."



Letters between Georges Lemaître and William Smart reveal that there was no conspiracy behind the removal of paragraphs from a translated paper.

of his earlier manuscript. Smart asked Lemaître whether he would allow his 1927 paper to be reprinted in the *Monthly Notices*, because the Royal Astronomical Council felt that the paper was not as well known as it should be. The most important paragraph in the letter reads:

“Briefly — if the Soc. Scientifique de Bruxells is also willing to give its permission — we should prefer the paper translated into English. Also, if you have any further additions etc on the subject, we would glad[ly] print these too. I suppose that if there were additions a note could be inserted to the effect that §§1–n are substantially from the Brussels paper + the remainder is new (or something more elegant). Personally and also on behalf of the Society I hope that you will be able to do this.”

In my view, Smart’s letter seems innocent; there is no suggestion of extra editing or censorship. Still, Block inferred from it hints of a conspiracy³. Block proposed that the handwritten “§§1–n” should be read as “§§1–72”, indicating freedom to translate only paragraphs 1–72 of his paper; where paragraph 73 was Lemaître’s equation determining the value of the Hubble constant. Block also claimed that Lemaître was effectively being told by Smart that Hubble’s observational result of 1929 is “something more elegant”.

I was not convinced by these claims: “n” makes more sense as a simple place-holder for the end of Lemaître’s article, and the shape of the alleged number “2” does not match the same numeral that appears later. Nevertheless, the mysteries of who translated the paper and who deleted the paragraphs remained unresolved.

CRUCIAL EVIDENCE

To definitively answer these questions, I obtained permission from Royal Astronomical Society librarian Peter Hingley, and Bob Carswell, the editor-in-chief of *Monthly Notices of the Royal Astronomical Society*, to scrutinize all of the Royal Astronomical Society Council’s minutes and the entire surviving correspondence from 1931. Eventually, I discovered two crucial documents.

First, in the minutes of the council from 13 February 1931, it is reported that¹⁰: “On the motion of Dr. Jackson it was resolved that the Abbé Lemaître be asked if he would allow his paper “Un Univers Homogène de Masse Constante et de Rayon Croissant,” or an English translation thereof, to be published in the *Monthly Notices*.” This, of course, was the decision mentioned in Smart’s letter to Lemaître.

Second, I found Lemaître’s response to Smart’s letter, dated 9 March 1931. The letter reads:

“Dear Dr. Smart

I highly appreciate the honour for me and for our society to have my 1927 paper reprinted by the Royal Astronomical Society. I send you a translation of the paper. I did not find advisable to reprint the provisional discussion of radial velocities which is clearly of no actual interest, and also the geometrical note, which could be replaced by a small bibliography of ancient and new papers on the subject. I join a french text with indication of the passages omitted in the translation. I made this translation as exact as I can, but I would be very glad if some of yours would be kind enough to read it and correct my english which I am afraid is rather rough. No formula is changed, and even the final suggestion which is not confirmed by recent work of mine has not been modified. I did not write again the table which may be printed from the french text.

As regards to addition on the subject, I just obtained the equations of the expanding universe by a new method which makes clear the influence of the condensations and the possible causes of the expansion. I would be very glad to have them presented to your society as a separate paper.

I would like very much to become a fellow of your society and would appreciate to be presented by Prof. Eddington and you.

If Prof. Eddington has yet a reprint of his May paper in M.N. I



Edwin Hubble at work in California’s Mount Wilson Observatory in 1937.

would be very glad to receive it.

Will you be kind enough to present my best regards to professor Eddington.”

This clearly ends speculation about who translated the paper and who deleted the paragraphs — Georges Lemaître did both himself.

Lemaître’s letter also provides an insight into the scientific psychology of (some of) the scientists of the 1920s. Lemaître was not at all obsessed with establishing priority for his original discovery. Given that Hubble’s results had been published in 1929, Lemaître saw no point in repeating his own more tentative earlier findings in 1931. Rather, he preferred to move forward and to publish his new paper, ‘The expanding Universe’, which he did later that year¹¹. Lemaître’s request to join the Royal Astronomical Society, at Smart’s invitation, was eventually granted; he was elected as an associate on 12 May 1939. ■

Mario Livio is at the Space Telescope Science Institute, Baltimore, Maryland 21218, USA.
e-mail: mlivio@stsci.edu

1. Nussbaumer, H. & Bieri, L. preprint at <http://arxiv.org/abs/1107.2281> (2011).
2. Van den Bergh, S. preprint at <http://arxiv.org/abs/1106.1195> (2011).
3. Block, D. preprint at <http://arxiv.org/abs/1106.3928> (2011).
4. Lemaître, G. *Ann. Soc. Sci. Brux.* **A 47**, 49–59 (1927).
5. Stromberg, G. *Astrophys. J.* **61**, 353–362 (1925).
6. Hubble, E. P. *Astrophys. J.* **64**, 321–369 (1926).
7. Hubble, E. P. *Proc. Natl Acad. Sci. USA* **15**, 168–173 (1929).
8. Lemaître, G. *Mon. Not. R. Astron. Soc.* **91**, 483–490 (1931).
9. Reich, E. S. *Nature* <http://dx.doi.org/10.1038/news.2011.385> (2011).
10. Royal Astronomical Society papers 2, Minutes of Council **12**, 160, 165, 166 (1931).
11. Lemaître, G. *Mon. Not. R. Astron. Soc.* **91**, 490–501 (1931).



Salvator Mundi was authenticated as a Leonardo by details such as the depiction of refraction in the orb.

ART HISTORY

Sight and salvation

Martin Kemp sifts the evidence that Leonardo da Vinci painted the newly emerged *Salvator Mundi*.

New paintings by Leonardo da Vinci do not come along every week, although the volume of messages in my inbox from owners of putative Leonardos might suggest to the contrary. The last painting that gained general acceptance as a Leonardo is the Benois Madonna (*Madonna with a Flower*), which made its public debut in 1909.

Now we have the newly emerged *Salvator Mundi* (1502–08) — Christ depicted as

saviour of the world. It is a highly traditional subject that even Leonardo had to lay out in a conventional manner. A fully frontal Christ, who looks unrelentingly at us, blesses with one hand and holds an orb in the other. The painting made its public debut on 9 November at the National Gallery in London, as part of the exhibition *Leonardo da Vinci: Painter at the Court of Milan* — the greatest gathering of Leonardo's paintings ever organized.

Leonardo da Vinci: Painter at the Court of Milan
National Gallery,
London.
Until 5 February 2012.

There are numerous copies and versions of *Salvator Mundi*, as there were of many of Leonardo's small-scale paintings. How do we know if one of them is

actually by Leonardo? There are two main resources that can be used: traditional art-historical arguments and newer techniques of scientific examination.

The standard art-historical arguments centre around connoisseurship — the validation of the attribution by an expert's eye. Although connoisseurship still has a role to play, and many experts depend on it, it involves subjective criteria that should long have been superseded as the key tool of attribution.

Other art-historical evidence can help. Two drawings by Leonardo in the Royal Collection at Windsor, UK, show the draperies of Christ. There is also a record that Leonardo's pupil, Gian Giacomo Caprotti (called Salaì), owned a "Christ in the Manner of God the Father". Salaì, to judge from documents of 1525 that record his property after his death, had managed to lay his hands on a central stock of Leonardo paintings, including the *Mona Lisa*. What happened to *Salvator Mundi* after that is unclear, although it did appear, in the seventeenth century, in records of the English collections of King Charles I, King Charles II and the Duke of Buckingham.

We can now supplement the art-historical arguments with the scientific. The evidence is of two kinds.

The first involves the technical examination of the painting, as undertaken during its recent conservation. It revealed that once, Christ's thumb was more upright, suggesting that this painting is not a straight copy. More importantly, examination by infrared reflectography, which involves bouncing infrared light off the white priming of the panel, has revealed characteristic signs of Leonardo's idiosyncratic technique. Particularly significant are clear signs of the painter pressing the edge of the palm of his hand into the wet paint, as can be seen above Christ's left eye. This was one of the techniques Leonardo used to create the soft, elusive effects for which he was renowned.

The second variety of 'scientific' evidence is particular to Leonardo. He insisted that painting is a science — it relies on a systematic body of knowledge based on a deep scrutiny of cause and effect in nature. He saw painting as "the sole imitator of all the manifest works of nature ... which with philosophical and subtle speculation considers all manner of forms ... all of which are enveloped in light and shade". For any painting to be recognized as a Leonardo, it has to bear witness to such mighty ambitions. The *Salvator Mundi* does on two main optical counts.

Leonardo was always interested in the

LEONARDO DA VINCI/PRIVATE COLLECTION © 2011 SALVATOR MUNDI LLC/PHOTO: T. NIGHSWANDER/IMAGINGART

functioning of sight, and in a manuscript largely devoted to the optics of the eye, written around 1507, he turned his attention to the issue of focus — what a photographer would call depth of field. He realized that something too close to the eye is not seen clearly, although he did not have any conception that this resulted from the focal range of the lens. He also stressed the loss of clarity when something became more distant. As he argued that images were received in the eye across a receptive surface, not at a single point, he maintained that the eye does not perfectly “know the edge of any body”.

In *Salvator Mundi* he plays with these depth-of-field problems. None of the contours is absolutely sharp, but the blessing hand and the tips of the fingers cradling the orb are discernibly clearer than the features of Christ's face. The rapid loss of clarity in depth serves to give space to what would otherwise be quite a flat image.

The other optical effect is unique to this painting, both in Leonardo's work and in the Renaissance more generally. The orb is not the standard globe of the world. It is translucent and glistens internally with little points of light. These are not the spherical bubbles found in glass, but are the kind of cavity inclusions (small gaps) that appear in some specimens of rock crystal and calcite. Leonardo, we know, was considered an expert in such semi-precious materials. It seems that he observed the double refraction produced by calcite. The heel of Christ's hand exhibits two distinct contours, not in this case due to a change of mind.

The crystal orb is not simply a visual tour de force. It reworks the meaning of the painting. In Ptolemaic cosmology, the stars were embedded in the fixed crystalline sphere of the heavens. The saviour of the world is also the saviour of the cosmos. Leonardo characterized God, who created the whole system, as the ‘Prime Mover’, who stands outside the sphere of the fixed stars and who set everything in motion at the moment of creation.

Leonardo has taken a stock subject and recast it. The soft, out-of-focus effects for Christ's head, endowing him with an ambiguous expression, invite the viewer to speculate on the supreme mystery of his flesh-and-blood presence on Earth. Leonardo also signals, through the crystalline sphere, that the domain of Christ's Father extends to the whole of the cosmos.

However skilled Leonardo's followers and imitators might have been, none of them reached out into such realms of “philosophical and subtle speculation”. We cannot reasonably doubt that here, we are in the presence of the painter from Vinci. ■

Martin Kemp is emeritus professor of art history at the University of Oxford, UK. *Salvator Mundi* features in his books *Christ to Coke* (2011) and *Leonardo* (2004).

Books in brief



The Quantum Universe: Everything That Can Happen Does Happen

Brian Cox and Jeff Forshaw ALLEN LANE 272 pp. £20 (2011)

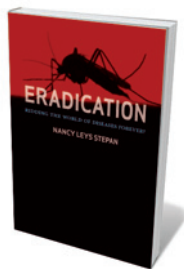
If quantum entanglement leaves you tongue-tied or you burn to know what fills ‘empty’ space, this offering from Brian Cox and Jeff Forshaw is a solid introduction to the “inescapable strangeness” of the subatomic world. Particle physicist and presenter Cox and theoretical physicist Forshaw nip through the territory with brio, unveiling the quantum cornucopia with clarity and concision — from the double-slit experiment, the wave-particle phenomenon and the key principles and constants, to the illusion of movement, the uneven spin of electrons and the death of stars.



Baby-Making: What the New Reproductive Treatments Mean for Families and Society

Bart Fauser and Paul Devroey OXFORD UNIVERSITY PRESS 289 pp. £16.99 (2011)

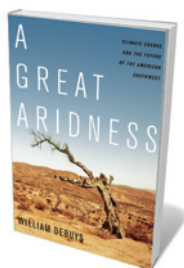
One in 25 European infants now gets its start in the test tube. But taking the sex out of conception has spawned vast scientific, ethical and social complexities. In this succinct overview, fertility experts Bart Fauser and Paul Devroey cover techniques from cryopreservation of embryos to egg donation. They explore baby ‘design’, parental demands, infertility, genetic issues and the social, ethical and scientific limits of assisted reproduction.



Eradication: Ridding the World of Diseases Forever?

Nancy Leys Stepan REAKTION BOOKS 312 pp. £25 (2011)

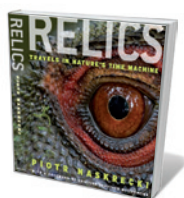
The idea of vanquishing diseases globally, one by one, has been contested ever since it emerged in the twentieth century. Historian Nancy Leys Stepan's book ranges from imperialist politics to medical technology. She probes the role of the World Health Organization and the Rockefeller Foundation, the efforts of “arch-eradicationist” Fred Lowe Soper to wipe out malaria, and other campaigns such as that against Guinea worm disease. Success, she shows, is rarely absolute: the 1980 eradication of smallpox was a triumph that now, with the threat of bioterrorism, teeters into uncertainty.



A Great Aridness: Climate Change and the Future of the American Southwest

William deBuys OXFORD UNIVERSITY PRESS 384 pp. \$27.95 (2011)

The bone-dry American Southwest is a trainwreck waiting to happen, says writer William deBuys. A swelling population with water-guzzling habits, combined with the impacts of climate change, threaten the balance of the region's vast interlocking ecosystems. Drawing on the work of climatologists and other scientists, deBuys's analysis of the eco-crisis — rising temperatures, wildfires, water shortages, disappearing wildlife — is a reasoned warning to heavily populated arid regions round the world.



Relics: Travels in Nature's Time Machine

Piotr Naskrecki UNIVERSITY OF CHICAGO PRESS 384 pp. £26 (2011)

Conservation biologist and photographer Piotr Naskrecki delves into ‘deep time’ in this gorgeously illustrated paean to flora and fauna that have changed little from their fossil forebears. Through his lens, we get a hint of prehistoric times from ecosystems such as the Papuan highlands or New Zealand's fern forests. His photos of ancient subjects — including Indonesian forest dragons (*Hypsilurus dilophus*) and the sagebrush cricket (*Cyphoderris strepitans*) of Wyoming — are a call to conserve these venerable survivors.

IN RETROSPECT

The art of influence

Thomas Dietz reassesses Robert Cialdini's revolutionary treatise on the science of decision-making.

We make decisions based on narrow self-interest, calculated benefits, costs and risks. Or so claimed economist Adam Smith, whose 'rational actor model', from his 1776 opus *The Wealth of Nations*, has long dominated thinking in economics and social science. By the late twentieth century, Smith's view had been applied to every domain of human decision-making, from marriage to international negotiations. But a growing body of evidence began to indicate that the model was often misleading.

Robert Cialdini's 1984 book *Influence* moved scientific thinking forward by showing that decision-making was messy. Whereas Smith's concept was logical and easy to describe mathematically, Cialdini grasped that decision-making is so complex that no single model can capture it. Analysing the literature on how we make decisions, he offered a picture that is more than a survey of specific research results, yet not quite an overarching unitary theory. The resulting treatise led Cialdini to become one of the most cited social psychologists; his books have sold more than two million copies.

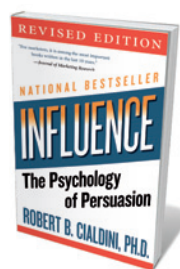
Cialdini saw that Smith's model of rational decision-makers, immune to any influence other than information, was simplistic. In reality, we are bombarded by efforts to persuade us to make decisions that serve the persuader's interests. If we make decisions based on pure logic, why would companies spend billions on advertising? Why would salespeople do more than list prices and options? Why would politicians focus so intently on symbolic messages?

These groups use our decision-making vulnerabilities as a fulcrum. Most people, most of the time, use shortcuts to choose between options — including being swayed by emotions, symbols and norms. This saves time and effort, and Cialdini argued that the shortcuts usually serve us well. But they can lead us badly astray, especially when manipulated by unscrupulous agents. *Influence* was intended to be a remedy against such manipulation — which Cialdini called the "real treachery, and the thing we cannot tolerate".

Cialdini articulated six "weapons of influence" — shortcuts that can be manipulated to shape our decisions. These comprise our tendency to reciprocate small favours with

larger ones; to stick with commitments and be consistent; to believe what others believe; to go along with what is suggested by those we like; to believe in authority; and to value things that seem scarce. For example, salespeople often start with a desirable offer to get a commitment, then throw in something for free to invoke reciprocity. They will look for clues about the hobbies and interests of customers (to seem likeable) and emphasize the popularity and scarcity of a product to make it seem more desirable.

Cialdini's use of the terms "weapons" and "treachery" reflects his belief that people and communities can suffer when salespeople, multinational corporations or governments work to shape our choices to their advantage. The power and authenticity of Cialdini's book comes in large part from the fieldwork that complements the experimental evidence. He engaged with salespeople, recruiters,



Influence: The Psychology of Persuasion
ROBERT B. CIALDINI
First published 1984.
Revised edition Harper:
2006. 336 pp. \$17.99,
£10.99



Perceived scarcity drives people to buy.

marketing people, staff at consumer agencies and others whose jobs are to persuade us or to uncover abuses of persuasion. He even went on sales-training courses to learn how our vulnerabilities can be exploited. These anecdotes make for an engaging read while underlining that persuasive ability is a form of power, often used in destructive ways.

Influence made the nuances of decision-making accessible to a broad audience, and research in that field has since flourished. Neologisms such as 'behavioural economics' and 'neuro-economics' attest to the growing acceptance that the rational actor model is insufficient to describe how we make choices. Four Nobel prizes in economics (to Herbert Simon, Daniel Kahneman, Vernon Smith and Elinor Ostrom, Smith being the sole economist) have been awarded for work that, similar to Cialdini's, pushes beyond that model.

Mainstream policy analysis still relies heavily on the assumption of a rational decision-maker, but social psychology is starting to affect how policies are designed. In the 2008 book *Nudge* (very much a descendant of *Influence*), authors Richard Thaler and Cass Sunstein argue that insights from the social sciences — such as our strong tendency to choose the default option no matter what it is — can be used to encourage better decisions.

Thaler and Sunstein call for "libertarian paternalism", an approach in which everyone should be free to make choices, but which uses the principles of social psychology to push us towards options that are in our best interests. Their optimistic tone contrasts with Cialdini's concern with the effects of "compliance professionals", who design sales and marketing campaigns and who manipulate us against our own interests. But the scientific core is much the same.

Social psychologists are showing that, to address environmental problems — including huge challenges such as climate change — we will have to acknowledge that humanity does not always behave rationally. This is easy to see: households and organizations waste far more energy and emit far more greenhouse gases than they would if they were strictly rational, for example.

Understanding how individuals and organizations make decisions can suggest ways to close this energy-efficiency gap and significantly reduce emissions, as well as how to handle other environmental issues effectively. The influence of *Influence* will continue as we incorporate ideas that complement the traditional economic view into science and into policy approaches. Cialdini's legacy is a key starting point for many of us who work on these issues. ■

Thomas Dietz is professor of sociology and of environmental science and policy at Michigan State University, East Lansing, Michigan 48824, USA.

Correspondence

Nobels: Toll pioneers deserve recognition

Immunologists are delighted that the field of innate immunity has been recognized by this year's Nobel Prize in Physiology or Medicine. However, we believe that the Nobel Committee should also have acknowledged the seminal contributions of immunologists Charles A. Janeway Jr (1943–2003) and Ruslan Medzhitov.

Janeway laid out the major theory that unifies the principles of innate and adaptive immunity (C. A. Janeway *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13; 1989), later experimentally verified with Medzhitov and then by many others. He recognized that antigen alone is insufficient to elicit an adaptive immune response, and postulated that both primitive and higher animals have specialized 'pattern-recognition receptors' that induce the innate immune response when activated by a particular class of conserved microbial products ('pathogen-associated molecular patterns').

The most important tenet of his theory was the connection between innate immune signalling and initiation of the adaptive immune response through enhancement of antigen processing and presentation, induction of co-stimulatory signals and cytokine release.

Medzhitov and Janeway subsequently cloned a human 'Toll-like' receptor and showed that it activated signalling pathways that induce adaptive immunity (R. Medzhitov *et al.* *Nature* **388**, 394; 1997). This remarkable demonstration also provided a framework for interpreting the significance of Toll-like receptors and their ligands for the immune response.

The innate–adaptive connection is now a fundamental principle in

immunology. We believe that the work of Janeway and Medzhitov was a Nobel-standard breakthrough for immunology.

James P. Allison *Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, USA.*

Christophe Benoist *Harvard Medical School, Boston, Massachusetts, USA.*

Alexander V. Chervonsky* *University of Chicago, Illinois, USA. achervon@bsd.uchicago.edu*

**On behalf of 23 co-signatories listed alphabetically at <http://dx.doi.org/10.1038/479178a>.*

Nobels: maintaining Israel's record

Israeli science has done well in the past decade of Nobel prize awards, with five nominations to three leading institutions — or 0.77 laureates per million people (compared with 0.2 for the United States, for example). This is despite the country's tiny population and the fact that it is the world's leading exporter of brainpower. But more investment in education and research will be necessary to maintain this impressive record.

One-quarter of Israel's academic scholars work at leading academic institutions in the United States — five times more than from any other nation apart from Canada (just over 12%; see go.nature.com/xen9nws).

Since 1973, the proportion of scientists in the population and the government's investment in academia relative to gross domestic product have both dropped by more than half, putting Israel below the average for countries within the Organisation for Economic Co-operation and Development (see go.nature.com/vozwgh).

Another problem is that the ultra-orthodox education system — which is responsible for more than one-quarter of Israel's first-graders (from age six) — does not

teach science or mathematics.

But there is hope for the future. As shown by the ongoing demonstrations for social justice, one of which was estimated to involve 7% of Israel's Jewish population, activism is already leading to short- and long-term changes, including a drive for more investment at all levels of education.

Ilan Samish *Weizmann Institute of Science, Rehovot, Israel. ilan.samish@weizmann.ac.il*

Nobels: fundamental biology misses out

In 1895, Alfred Nobel bequeathed much of his immense fortune to create the five original Nobel prizes — for literature, physiology or medicine, physics, chemistry, and peace. The award categories were expanded to include economics in 1968, following a large donation to the Nobel Foundation on behalf of the Swedish central bank, by the creation of the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel.

The symbolic recognition of the work of Nobel laureates increases awareness of scientific research in the eyes of the public. It is a continuing shame that fundamental biology — and not just its application in medicine — lacks such a patron.

P. William Hughes *Carleton University, Ottawa, Canada. whughes@connect.carleton.ca*

Tropical forests: still vital when degraded

Luke Gibson *et al.* argue that tropical conservation efforts should focus on primary and selectively logged forests, because these are the most diverse (*Nature* **478**, 378–381; 2011). However, working in degraded habitats can also provide important conservation opportunities.

Worldwide, around 500,000 hectares of logged and secondary forests are converted every year to plantations such as oil palm, rubber, *Acacia* and *Eucalyptus*, reducing both their current biodiversity value and their capacity for regeneration. Because of the revenue they generate, conserving such forests may be more economically viable than the total protection of pristine forests.

Disturbed forests include important 'matrix habitats' between areas of intact forest that can buffer against the impact of habitat fragmentation on reserves. Species occupancy of protected areas is more strongly associated with the condition of the matrix than with either reserve size or isolation (L. R. Prugh *et al.* *Proc. Natl Acad. Sci. USA* **105**, 20770–20775; 2008). Such forests also provide crucial habitat for wide-ranging species.

In many parts of the tropics, primary forests are confined to upland areas or to poor soils, and are biogeographically restricted. It may be more productive to focus on restoring degraded, but endangered, lowland habitats than on protecting yet more upland forest.

Another consideration is that hunting in tropical reserves significantly threatens their long-term viability. Protection of wildlife may be better achieved through working with timber or oil concessions, or in buffer-zone tourist reserves, than with cash-strapped (or otherwise compromised) nature-reserve agencies.

Tropical reserves cover roughly 11% of the tropical forest biome, whereas 40% is deforested, 20% is logged and an unknown but significant proportion of the remainder is severely degraded (S. J. Wright *Annals New York Acad. Sci.* **1195**, 1–27; 2010). Such figures alone make it necessary to consider the protection of biodiversity in

degraded tropical habitats.

Rhett D. Harrison

*Xishunagbanna Tropical
Botanical Garden, Chinese
Academy of Science, Menglung,
Yunnan, China.
rharrison@xtbg.org.cn*

Tropical forests: try holistic conservation

Sustaining tropical biodiversity in the long term requires more than just protection of remaining primary forests (L. Gibson *et al.* *Nature* **478**, 378–381; 2011). It is essential to encourage land-use strategies that increase the potential conservation value of habitats modified by humans. Substantial tropical biodiversity can also thrive in such human-dominated ecosystems.

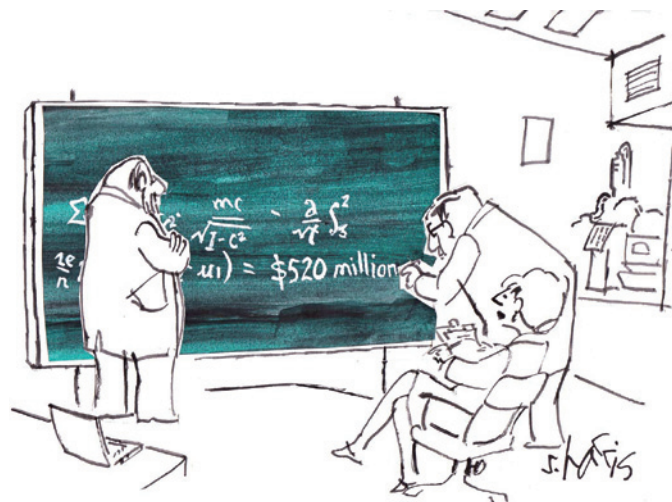
Luke Gibson and colleagues highlight the importance of disappearing forests by comparing biodiversity in tropical primary forests with corresponding human-modified habitats. But clinging to the ecological ideal excludes equally important comparisons to ecological dystopias. For example, when I re-analysed the data to include a 'degraded' reference state with poor biodiversity (see www.stanford.edu/~cdm/Pubs/SOM), it revealed that human-dominated ecosystems sustain at least a medium amount of biodiversity. This is substantial, considering that such ecosystems constitute roughly 85% of the tropics (E. C. Ellis *et al.* *Global Ecol. Biogeogr.* **19**, 589–606; 2010).

Rather than championing conservation of primary forest over land-use strategies favouring conservation, we must develop a holistic conservation model that includes human-dominated ecosystems.

Chase D. Mendenhall *Center for Conservation Biology, Stanford University, California, USA.
cdm@stanford.edu*

Tropical forests: include Congo basin

Luke Gibson and co-workers provide an important global assessment of the impact of disturbance and land conversion



on biodiversity in tropical forests (*Nature* **478**, 378–381; 2011). But their meta-analysis of 138 studies overlooks the Congo basin, the second-largest continuous area of rainforest in the world; moreover, only 12 studies are located in Africa. This omission is not the fault of the authors, but is symptomatic of the lack of recent and accessible legacy data for this region.

The Democratic Republic of Congo (DRC), which contains 98 million hectares of rainforest (60% of the Congo basin forest), is also poorly represented in studies beyond biodiversity. Take, for example, a study that estimates the carbon sink in the world's forests on the basis of inventory data and long-term ecosystem studies (Y. Pan *et al.* *Science* **333**, 988–993; 2011). The study's estimate for carbon stocks in intact tropical forest across Africa is based on a network of 79 monitoring sites, yet only ten of these are in the DRC, all in the same forest reserve in the northeastern Ituri province.

Large scientific, logistic and training efforts are needed to establish permanent monitoring sites in the Congo basin's tropical forests and woodlands, and to connect those sites to global networks.

Countries such as the DRC must also identify and monitor forest biodiversity and carbon stocks within the framework of international climate and conservation policies, such as the United Nations REDD programme (for 'reducing emissions from deforestation

and forest degradation'). This will help to preserve the Congo basin forest, where deforestation so far remains modest compared with that in the Amazon region and in southeast Asia.

Hans Verbeeck, Pascal Boeckx, Kathy Stepe *Ghent University, Belgium.*

hans.verbeeck@ugent.be

Clarifying the use of 'prepubescent'

By implying that 11- and 12-year-old girls are "prepubescent", you play into the hands of those who oppose vaccinating young girls against human papilloma virus (HPV) to prevent cervical cancer (*Nature* **477**, 369; 2011).

Although the average age of first menstruation (menarche) was 12.8 years in the United States in 1968, this is becoming steadily younger (see, for example, S. E. Anderson and A. Must *J. Pediatr.* **147**, 753–760; 2005). Also, as menarche comes quite late in a girl's progress towards full adult height and reproductive capacity, girls should no longer be described as prepubescent at ages 11 and 12.

A case could be made for scheduling vaccination against HPV at an even younger age — at the start of breast development, or around age 10, before a girl is likely to become sexually active.

Evelyn J. Bowers *Ball State University, Muncie, Indiana, USA.
00ejbowers@bsu.edu*

More than one fund for US entrepreneurs

There are more funding opportunities for entrepreneurial US scientists than Steve Blank conveys (*Nature* **477**, 133; 2011). These are especially pertinent in a year when the National Science Foundation (NSF) is seeking congressional reauthorization for its Small Business Innovation Research (SBIR) Program.

Although the NSF's US\$5-million Innovation Corps (I-Corps) initiative is welcome, the SBIR Program has available funds of about \$150 million a year. Overall, SBIR programmes at the NSF, NASA and the US government departments of energy, defence and commerce provide \$2 billion in funding a year. Since their inception in 1982, they have enabled a whole generation of scientists to become entrepreneurs.

The I-Corps sends funding directly to universities rather than to small businesses. However, many SBIR-funded small businesses were originally university spin-offs, and the complementary Small Business Technology Transfer (STTR) programme already targets partnerships between universities and small businesses.

Probably the most innovative element of the I-Corps programme is its partial funding by private foundations. If this could be increased, it would help to alleviate pressure on federal funding of the SBIR and STTR programmes.

SBIR funding was significant to the success of both of the science-entrepreneurial spin-offs in which I was involved after leaving Harvard University (see www.aer.com and www.cri-inc.com). These companies have since gone public, having generated products for worldwide use in applications ranging from biomedical imaging to environmental control.

Peter Foukal Nahant, *Massachusetts, USA.
pvfoukal@comcast.net*

The path to sociality

A comparative analysis traces the trajectory of change in social organization among primates and establishes a firm foundation for modelling the origins of social complexity. [SEE LETTER P.219](#)

JOAN B. SILK

Recent discoveries related to human origins — from new fossils to ancient DNA — have stirred intense interest from scientific journals and the popular media. But perhaps more intriguing are the intimate lives of our distant ancestors. Although the steamy details may be shrouded in the mists of time, some hints about the sequence of events that led to the evolution of human social systems are emerging. The latest evidence comes from Shultz *et al.*¹, who on page 219 of this issue trace the evolution of complex sociality within the order Primates. Their data provide a strong foundation for modelling the origins of hominid mating systems by constraining the range of likely trajectories of social change.

Decades of hard work by hundreds of field researchers have provided a rich body of knowledge about the social organization of non-human primates². Like many mammals, members of a few primate species lead a largely solitary life as adults, meeting only to mate or exchange calls at the borders of their territories (for instance, Coquerel's dwarf lemur, *Mirza coquereli*). By contrast, other primates are found in groups consisting of one adult male and multiple adult females (gorillas), or form large groups consisting of several males and females (baboons), or pair up to form lasting bonds and live in family groups (gibbons and humans).

Shultz *et al.*¹ set out to assess the effects of phylogenetic history on social organization and to examine the pattern of change in such organization across time. For this, they combined information about the social organization of 217 living primate species with information about the genetic relationships among species within the order. They conclude that social organization has a strong phylogenetic component, because it tends to be more similar among closely related species than would be expected by chance.

The authors also evaluate the patterns of change from one form of social organization to another. It seems logical that solitary social systems would give rise to stable bonds between pairs, as temporary rendezvous between males and females for mating are lengthened into enduring associations. It also makes sense



Figure 1 | Living together. Shultz and colleagues' study¹ throws light on developments that led to the stable communities in which chimpanzees exist today.

that there would be a steady progression from small groups to larger, more complex ones. But apparently this is not what actually happened. Instead, Shultz and colleagues' phylogenetic records reveal a set of largely unilateral transitions, some 52 million years ago, from solitary life to loose aggregations of multiple males and females. From there, more stable multi-male, multi-female groupings arose. And from the stable groups, both pair-bonded and one-male harems emerged roughly 16 million years ago.

The existence of a strong phylogenetic signal spells trouble for socioecological models that aim to explain the evolution of primate social organization. The models hypothesize that food distribution shapes competitive regimes, and that these, in turn, shape dispersal patterns and the nature of relationships within groups^{3–5}. These models generally assume that phylogeny does not impose notable constraints on social organization, and that changes from one form of social organization to another are all equally likely. But there is a growing realization that history does have a role^{6–8}, and the new results¹ strengthen that view.

Shultz and co-authors do not explore the

selective forces that favoured the transition from one stage of social organization to the next in particular primate lineages. However, they show that the initial shift to sociality coincided with a transition from a nocturnal lifestyle, in which being inconspicuous is an important protective strategy against predators, to a diurnal lifestyle, in which there is safety in numbers.

Their data also provide insight into sex-biased dispersal patterns. Dispersal is a way to prevent inbreeding, and in many primate species, including chimpanzees, members of one sex move away from their natal group, whereas members of the opposite sex stay behind throughout their lives. In these species, sex-biased dispersal patterns allow the formation of cooperative relationships among members of the resident sex⁹. Shultz *et al.* report that sex-biased dispersal patterns arose after the transition to multi-male, multi-female groups — not before. This suggests that the benefits of cooperating with resident kin did not drive the evolution of sociality in primates. Instead, stable residence patterns provide opportunities for forming cooperative relationships with group members, particularly with close kin.

The direction of the transition from stable groups with multiple males and females to pair-bonded groups is of particular interest for at least two reasons. First, it challenges assumptions about the nature of social complexity. In the past few decades, the social-brain hypothesis has held sway in primatology^{10,11}. According to this hypothesis, the demands of living in large groups with a host of potential rivals favoured the evolution of greater cognitive abilities: the larger the group, the more complicated the social terrain and the greater the need for cognitive sophistication. But Shultz and colleagues' finding that there has not been a steady progression from small groups to larger ones suggests that social complexity is not a simple function of group size.

Second, these findings provide some insight into events that may have occurred within the human lineage. The last common ancestor of humans and chimpanzees lived about 5 million to 7 million years ago. We know nothing about

the social lives of these creatures, but we do know that modern chimpanzees live in stable communities consisting of multiple males and females (Fig. 1). The new results make it unlikely that the last common ancestor was pair-bonded, because this would imply an improbable transition from pair-bonds to larger groupings within the chimpanzee lineage. This also means that the shift to pair bonds must have occurred sometime after the divergence from this ancestor.

Shultz and colleagues' paper¹ is unlikely to attract the kind of fanfare that accompanies the discovery of a new hominid fossil or the sequencing of an ancient genome. But it is likely to play an essential part in modelling the evolution of sociality in primates, in interpreting the fossil record and in reconstructing the lives of our ancestors. ■

Joan B. Silk is in the Department of Anthropology and the Center for Society and

Genetics, University of California, Los Angeles, Los Angeles, California 90064, USA.
e-mail: jsilk@anthro.ucla.edu

1. Shultz, S., Opie, C. & Atkinson, Q. D. *Nature* **479**, 219–222 (2011).
2. Mitani, J., Call, J. C., Kappeler, P. M., Palombit, R. A. & Silk, J. B. (eds) *The Evolution of Primate Societies* (Univ. Chicago Press, in the press).
3. Sterck, E. H. M., Watts, D. P. & van Schaik, C. P. *Behav. Ecol. Sociobiol.* **41**, 291–309 (1997).
4. Isbell, L. A. & Young, T. P. *Behaviour* **139**, 177–202 (2002).
5. Koenig, A. *Int. J. Primatol.* **23**, 759–783 (2002).
6. Rendall, D. & di Fiore, A. *Evol. Anthropol.* **4**, 43–52 (1995).
7. Kappeler, P. M. & van Schaik, C. P. *Int. J. Primatol.* **23**, 707–740 (2002).
8. Thierry, B. *Evol. Anthropol.* **17**, 93–96 (2008).
9. Silk, J. B. *Phil. Trans. R. Soc. B* **364**, 3243–3254 (2009).
10. Dunbar, R. I. M. *Annu. Rev. Anthropol.* **32**, 163–181 (2003).
11. Cheney, D. L. & Seyfarth, R. M. *Baboon Metaphysics: The Evolution of a Social Mind* (Univ. Chicago Press, 2007).

PLANETARY SCIENCE

Ancient lunar dynamo

The differential rotation between the Moon's core and mantle may have powered the ancient lunar dynamo, either continuously over several hundred million years or intermittently after large impacts. SEE LETTERS P.212 & P.215

DOMINIQUE JAULT

Possible energy sources have long been sought for a dynamo that could have produced the magnetic field possessed by the Moon 4 billion years ago¹. In this issue, Dwyer *et al.*² and Le Bars *et al.*³ show how the ancient lunar dynamo, which acted in the Moon's fluid core, may have been mechanically driven.

The two teams describe fluid dynamos that could account for the magnetization of the lunar crust^{4,5}. Both groups investigate the fluid motions spawned by the fast spin of the Moon's fluid core past its solid mantle. Dwyer *et al.*² (page 212) consider that the misalignment between the rotational axes of the core and the mantle — which has persisted throughout the Moon's history — resulted from the almost perfectly spherical shape of the lunar core. Le Bars *et al.*³ (page 215) examine the sudden differential rotation between the core and the mantle that followed large impacts on the Moon's surface. Le Bars and colleagues' dynamo works only briefly, but for long enough to magnetize the basins formed after the impacts that ignited the dynamo process.

Dwyer *et al.* build on a finding⁶ that predated the Apollo programme: that the small ellipticity of the Moon's core surface means that the lunar core does not follow the mantle

in its precession. Inertial coupling, which results from centrifugal pressure acting on Earth's elliptical core–mantle boundary, causes Earth's core and mantle to precess in sync. This mechanism is not viable for slowly rotating bodies such as the Moon as it is today. According to Dwyer *et al.*, it has not been active for most of lunar history, during which

the lunar core's rotational axis has remained nearly normal to the ecliptic (the plane of Earth's orbit around the Sun): viewed from the core, the mantle rotates about an equatorial axis, which revolves about the core's rotational axis. Analyses based on lunar laser ranging — which involve measuring round-trip travel times of laser pulses between Earth and the Moon — have confirmed⁷ that, at present, the core rotational axis is normal to the ecliptic plane. Le Bars *et al.*³ investigate an alternative hypothesis, whereby the differential rotation was transient and organized about the Moon's polar axis.

The two studies^{2,3} give different reasons to expect differential motion between the liquid core and the solid mantle of the early Moon. From this perspective, it is advantageous to read the two articles side by side. Both exploit



Figure 1 | The Crisium lunar basin. Le Bars and colleagues' model³ for the ancient lunar dynamo explains magnetic anomalies of several impact basins, including Crisium.

recent work⁸ on flows in rotating ellipsoidal cavities: time-varying rotational motions of the cavity's outer solid boundary drive flows in the cavity with amplitude of the order of the rotational motions. Both the polar and the equatorial ellipticities of the Moon's core exceed the minimum value above which such motions arise: relative motions of the mantle about either a precessing equatorial axis or the polar axis lead to fluid instabilities in the Moon's core.

On the basis of their estimates for the amplitude of these mechanically forced motions in the ancient Moon, the authors^{2,3} argue that the motions powered a dynamo. Next, using scaling laws derived for convective dynamos⁹, they estimate the magnetic-field strength of the early Moon from the kinetic energy that was dissipated at the core–mantle boundary. To extrapolate the past dissipation rate from its current value⁷, Dwyer *et al.* and Le Bars *et al.* assume a turbulent friction mechanism, which implies that the dissipation scales as the cubic power of the differential rotation between core and mantle. Frictional coupling between the core and the mantle favours dynamo action by injecting kinetic energy at the boundary and also acts against it by limiting differential rotation.

The proposal of these two models should prompt further studies of mechanically driven dynamos. The ability of the predicted motions to produce a dynamo is likely but has yet to be proven, and the amplitude of the magnetic field that it would generate needs to be confirmed. Another question that needs to be addressed concerns the geometry of the magnetic fields. But the expectation is that these had to be on a predominantly large length-scale; otherwise, they would have been strongly attenuated between the core and the Moon's surface, and would not have significantly magnetized the crust.

Le Bars and colleagues' model nicely explains the central magnetic anomalies of a series of impact basins⁵, including Crisium (Fig. 1), but fails to account for the lack of anomalies inside the Imbrium and Orientale impact basins¹⁰, which formed rapidly after the basins possessing central magnetic anomalies. It is now crucial to understand why the impacts that formed these two large basins did not trigger a dynamo capable of leaving a magnetic imprint on the Moon's surface, as expected from the model of Le Bars and colleagues.

The two studies^{2,3} are thought-provoking and may be complementary. Le Bars *et al.* do not claim that their hypothesis accounts for all the evidence of an early lunar dynamo. Dwyer *et al.* present a paradoxical idea: that the Moon's low rotation rate, which has led to a small ellipticity of the core–mantle boundary, was once conducive to dynamo action. They link the onset of the dynamo to the decrease of the Moon's polar flattening as its rotation slowed. Future palaeomagnetic experiments on samples from very old lunar rocks will enable their theories to be tested. ■

Dominique Jault is currently at the *Institute of Geophysics, ETH Zürich, Switzerland, and in the Laboratoire ISTerre, Université Joseph-Fourier, CNRS, BP 53, 38041 Grenoble Cedex 9, France.*
e-mail: dominique.jault@obs.ujf-grenoble.fr

1. Stegman, D. R., Jellinek, A. M., Zatman, S. A., Baumgardner, J. R. & Richards, M. A. *Nature* **421**, 143–146 (2003).
2. Dwyer, C. A., Stevenson, D. J. & Nimmo, F. *Nature* **479**, 212–214 (2011).
3. Le Bars, M., Wiczeorek, M. A., Karatekin, Ö., Cébron, D.

- & Laneuville, M. *Nature* **479**, 215–218 (2011).
4. Garrick-Bethell, I., Weiss, B. P., Shuster, D. L. & Buz, J. *Science* **323**, 356–359 (2009).
5. Hood, L. L. *Icarus* **211**, 1109–1128 (2011).
6. Goldreich, P. J. *Geophys. Res.* **72**, 3135–3137 (1967).
7. Williams, J. G., Boggs, D. H., Yoder, C. F., Ratcliff, J. T. & Dickey, J. O. *J. Geophys. Res.* <http://dx.doi.org/10.1029/2000JE001396> (2001).
8. Cébron, D., Le Bars, M., Leontini, J., Maubert, P. & Le Gal, P. *Phys. Earth Planet. Inter.* **182**, 119–128 (2010).
9. Christensen, U. R., Holzwarth, V. & Reiners, A. *Nature* **457**, 167–169 (2009).
10. Mitchell, D. L. *et al.* *Icarus* **194**, 401–409 (2008).

ANIMAL BEHAVIOUR

Why promiscuity pays

A study reveals that female promiscuity in a songbird, the dark-eyed junco, is explained by the greater reproductive success of offspring sired by males outside social pairs compared with offspring born within pairs.

**LAWRENCE BELLAMY
& ANDREW POMIANKOWSKI**

Males and females of many species come together to tend and provide for their young. But it is no surprise that social bonds are often broken. Both sexes can be promiscuous, and the paternity of some offspring typically falls outside the social pair. For males, the benefits seem obvious — other parents take on the onerous task of raising their kids. But what benefits do females gain? A study by Gerlach and co-authors¹ in *Proceedings of the Royal Society B* uses long-term data on dark-eyed juncos (*Junco hyemalis*) to shed light on why females cheat on males.

Dark-eyed juncos (Fig. 1), and birds in general, are useful models for studying female promiscuity². Males and females form social pairs that work together to build a nest and rear offspring. However, this image of harmony is tainted by covert extra-pair matings that frequently lead to young. The natural assumption is that such sneak copulations allow females to increase the genetic quality of their offspring³ — especially for females stuck with genetically 'inferior' social partners.

The evidence for this hypothesis is, however, hardly convincing. A recent compilation of data from 55 bird species found that extra-pair offspring on average have barely any advantage over their within-pair siblings⁴. But the evidence provided by most comparisons of this sort is not strong, as the fitness components considered are restricted to survival of offspring to fledging, or size at independence. These are easy to assess, but are not necessarily closely linked to the mating success and reproduction of offspring.

Gerlach *et al.*¹ have used their long-term study to get a proper grip on fitness. They

monitored a population of juncos in western Virginia over 17 breeding seasons (and eight government grants). Each year, the team caught and took blood samples from all social pairs and their offspring (which were designated as F₁ offspring). Using DNA profiles, they identified whether the social male or an interloper had sired each chick. The following mating season, the authors identified F₁ offspring that returned to the breeding site as sexually mature adults, and calculated the birds' success in raising offspring in a subsequent generation (F₂), both as social mates and through extra-pair paternity.

In this way, Gerlach *et al.* obtained clear evidence that female promiscuity increased the lifetime fitness of offspring: on average, they observed that F₁ extra-pair offspring that returned to the population had about twice as many young as F₁ within-pair offspring. This advantage could not be accounted for by differences in survival rates, as the authors found that paternity did not affect the number of years F₁ offspring returned to breed.

The benefits of promiscuity accrued through reproductive success, but for different reasons in females and males. F₁ females that had extra-pair fathers were superior to those that had within-pair fathers because they had more offspring sired by their social partner. These F₁ females may have laid more eggs, been better mothers, or attracted superior paternal care from their social partners. By contrast, they did not have a greater number of young through extra-pair matings.

The success of F₁ males that had extra-pair fathers was largely accounted for by their greater ability to sire young outside their social partnership: they were more than twice as likely as F₁ males that had within-pair fathers to sire young in another pair's nest. But



Figure 1 | A model of promiscuity. Gerlach *et al.*¹ report that the offspring of female dark-eyed juncos have greater lifetime fitness when sired by a male outside the social pair, suggesting an evolutionary advantage for female promiscuity.

these F_1 extra-pair males did not benefit their social partners, as they had no influence on the number of offspring produced within their home nests. Nor were they any better or worse than their male counterparts born from social pairs at defending their social partner from the attentions of other males: the number of offspring lost to extra-pair paternity was the same for F_1 extra-pair and within-pair males.

The reproductive advantages observed by Gerlach *et al.* for the offspring of promiscuous matings tie in with several theoretical predictions. If females prefer sneak matings with attractive males, one would expect them to have more-attractive sons, which in turn will sire more extra-pair offspring. This is a variant of evolutionary biologist Ronald Fisher's theory of runaway sexual selection⁵. Although Fisher was focusing on the evolution of peacocks and other highly ornamented species that exhibit extreme polyandry (few males mating with many females), his logic also applies to female promiscuity within social pairing.

Gerlach *et al.* looked for direct support for runaway sexual selection in dark-eyed juncos, but found no evidence that F_1 extra-pair males were morphologically different from F_1 within-pair males. That said, the preference of junco females for promiscuous liaisons may be more strongly influenced by the song and display traits of males than by morphology, but such traits were not measured in the authors' study.

The increased production of offspring by extra-pair daughters also cannot be explained by runaway sexual selection, but it does suggest that there is a positive genetic covariance between male attraction and female offspring production. Such an advantage is predicted by 'good genes' models of sexual selection if genetic quality underlies both a male's capacity to attract extra-pair matings and a female's fecundity and parenting ability^{6,7}.

In Gerlach and colleagues' study¹, the additional offspring produced by F_1 extra-pair females were sired by the social partner, and did not come from extra-pair liaisons. How this is achieved and what purpose it might serve is rather mysterious. Perhaps F_1 extra-pair females pair with males that are in better condition. These males might themselves contribute to the higher production of offspring through enhanced male parental care, or simply be more successful in warding off the attentions of extra-pair males and so grab a greater proportion of paternity. Another possibility is that females invest more in offspring sired by these superior males, as occurs in other species⁸, and indeed in juncos⁹, because such offspring will have inherited their father's better condition. Such ideas need further investigation.

Gerlach and colleagues' fitness analysis shows that the value of promiscuity becomes clear only through multi-generational, direct measurements of viability and reproductive success from parents down to their grand-offspring. This is needed not only for the offspring of social pairs, but also for those derived from promiscuous copulations; measuring proxy components of fitness provides only half-answers. The authors' approach will therefore be a model for future studies in this field. ■

Lawrence Bellamy is in the Department of Genetics, Evolution and Environment, Darwin Building, University College London.

Andrew Pomiankowski is in the Department of Genetics, Evolution and Environment, and also at the Centre for Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London WC1E 6BT, UK.

e-mails: lawrence.bellamy@ucl.ac.uk; ucbpom@ucl.ac.uk



50 Years Ago

Computer Logic: The Functional Design of Digital Computers. By Dr. Ivan Flores — The book ... contains a very solid treatment of logical design; the binary system, adders, multipliers, control units, input and output devices, magnetic tape, and so on, all will be found described here, together with the usual ration of Boolean algebra and formal logic ... Only a small part of the book is concerned with programming, but this part I found unsatisfactory. It seems a mistake, now that modifier registers are the order of the day, to introduce the reader first to the old-time procedure for modifying instructions in the arithmetic unit. The last chapter, on programming a particular scientific problem, is a veritable museum piece ... It exhibits the most primitive form of machine language coding possible, in which even the conversion of relative addresses in sub-routines to absolute addresses must be done by the programmer himself before his programme is punched.

From *Nature* 11 November 1961

100 Years Ago

There is not the slightest doubt that birds and mammals are now being killed off much faster than they can breed. And it is always the largest and noblest forms of life that suffer most ... And the worst of it is that all this wanton destruction is not by any means confined to the ignorant or those who have been brought up to it. We have had our warnings. The great auk and the Labrador duck have both become extinct within living memory ... When wild life is squandered it does not go elsewhere, like squandered money; it cannot possibly be replaced by any substitute, as some inorganic resources are: it is simply an absolute dead loss, gone beyond even the hope of recall.

From *Nature* 9 November 1911

- Gerlach, N. M., McGlothlin, J. W., Parker, P. G. & Ketterson, E. D. *Proc. R. Soc. B* <http://dx.doi.org/10.1098/rspb.2011.1547> (2011).
- Birkhead, T. R. *Reprod. Fertil. Dev.* **7**, 755–775 (1995).
- Neff, B. D. & Pitcher, T. E. *Mol. Ecol.* **14**, 19–38 (2005).
- Akçay, E. & Roughgarden, J. *Evol. Ecol. Res.* **9**, 855–868 (2007).

- Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, 1930).
- Zahavi, A. J. *Theor. Biol.* **53**, 205–214 (1975).
- Iwasa, Y. & Pomiankowski, A. J. *Theor. Biol.* **200**, 97–109 (1999).
- Sheldon, B. C. *Trends Ecol. Evol.* **15**, 397–402 (2000).
- Grindstaff, J. L., Buerkle, C. A., Casto, J. M., Nolan, V. Jr & Ketterson, E. D. *Behav. Ecol. Sociobiol.* **50**, 312–316 (2001).

in senescent cells. The main known role of p16^{Ink4a} is to enforce cell-cycle arrest by activating the RB tumour-suppressor protein. There is also evidence that p16^{Ink4a} expression increases during ageing and is linked to age-related human disease⁸.

To define the role of senescent cells in ageing and tissue degeneration, Baker *et al.*³ specifically eliminated p16^{Ink4a}-expressing cells in mice. For this, they used an ingenious transgenic strategy⁹ in which specific cell groups could be killed by a drug-inducible caspase enzyme. Thus, the authors engineered transgenic mice in which toxicity could be selectively induced in p16^{Ink4a}-expressing cells.

Baker *et al.* intercrossed these animals with mice that express low levels of the BubR1 protein. This mitotic checkpoint protein suppresses untimely separation of sister chromosomes during cell division, thereby preventing the formation of cells with an abnormal number of chromosomes. In a largely p16^{Ink4a}-dependent manner, BubR1-insufficient mice suffer from age-related disorders prematurely¹⁰. The authors specifically killed p16^{Ink4a}-positive cells in the offspring of the intercrossed mice (Fig. 1).

In young animals with no signs of senescence, eliminating p16^{Ink4a}-positive cells profoundly delayed age-related deficits — such as hunchback and cataracts. Also, compared with control mice, their muscle fibres were thicker and their exercise fitness was significantly greater. What's more, long-term loss of p16^{Ink4a}-positive cells resulted in larger fat deposits, which indicate attenuation of ageing. As expected, this coincided with signs of clearance of senescent cells, such as reduced β -galactosidase activity and decreased expression across different tissues

AGEING

Old cells under attack

Age brings not just wisdom, but also, alas, many traits that to most of us are much less attractive. It now seems that, at least in mice, clearance of senescent cells delays some of the maladies associated with ageing. **SEE LETTER P.232**

DANIEL S. PEEPER

As we get older, most of us will at times wonder why ageing comes with degeneration, or, more pressingly, whether something can be done about it. Ageing causes progressive loss of function, decreased fertility and increased risk of mortality¹. This is relevant particularly to people living in modern societies. Animals in the wild often die young as a result of external hazards such as infection, predation and starvation¹. But in protected environments, intrinsic processes, mainly accumulating damage to DNA, impede tissue maintenance. Specifically, such damage triggers cellular senescence and cell death, both of which are thought to contribute to ageing². On page 232 of this issue, Baker *et al.*³ provide direct evidence for a relationship between cellular senescence and ageing.

Senescent cells can no longer divide; the cells are viable and metabolically active, but do not respond to cell-division stimuli. Such replicative senescence is caused by dysfunctional telomeres (DNA sequences at the ends of chromosomes). But other factors, including oncogenic events, can instigate senescence prematurely — that is, before telomeric abnormalities can take effect⁴.

Both telomere dysfunction and oncogenic alterations can induce the p53 and RB tumour-suppressor pathways⁵. Their activation is typically accompanied by the emergence of other hallmarks of senescence, such as increased activity of the enzyme β -galactosidase, as well as activation of a senescence-associated secretory program^{6,7}.

Accumulation of the p16^{Ink4a} protein, which is encoded by a tumour-suppressor gene that is frequently lost in cancer, is common

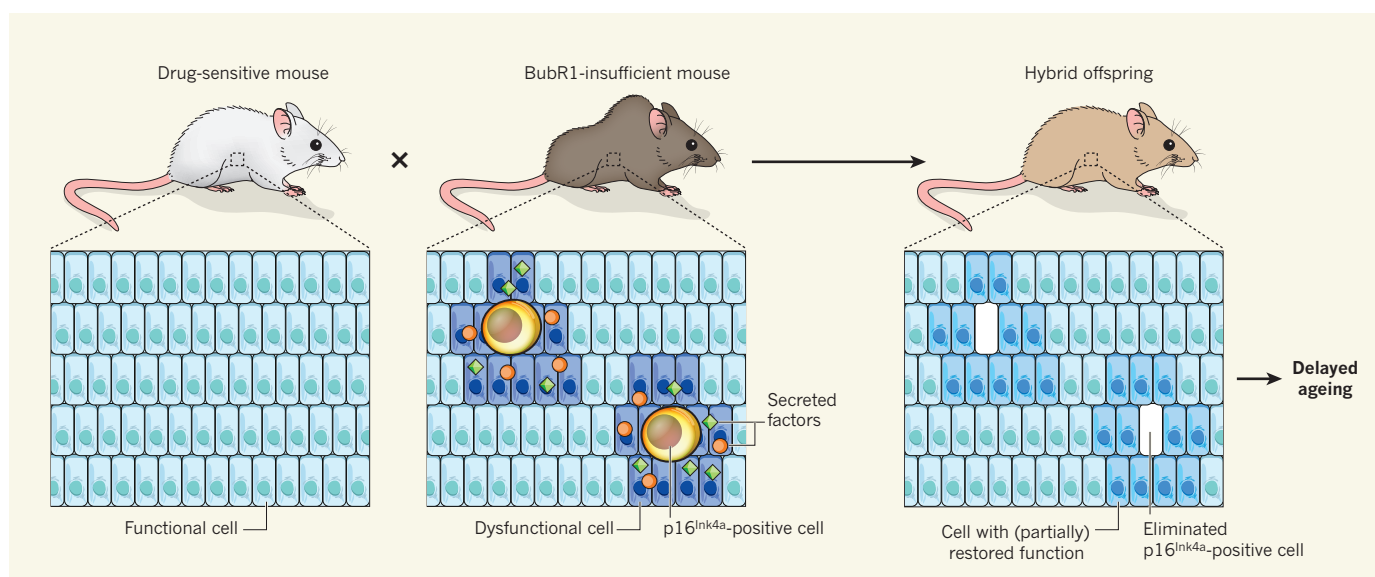


Figure 1 | Delaying ageing. Baker *et al.*³ generated mice in which p16^{Ink4a}-positive senescent cells could be selectively killed by drug treatment. They then crossed these otherwise normal mice with animals that express low levels of the BubR1 protein. The latter mice age prematurely, perhaps because

factors secreted by their accumulating p16^{Ink4a}-positive cells cause cellular and tissue dysfunction. When the authors eliminated p16^{Ink4a}-positive cells in the offspring of this intercross, the onset of age-associated deficits was delayed, causing improved cellular and tissue function.

of senescence-associated secreted factors, including IL-6 and matrix metalloproteinase enzymes.

These results suggest that continuous clearance of p16^{Ink4a}-positive senescent cells suspends the p16^{Ink4a}-dependent traits of premature ageing that are due to BubR1 deficiency. However, it would be more appealing if the ablation of senescent cells delayed features of ageing after the animals had begun to develop senescent traits. To see whether this was achievable, Baker *et al.* eliminated senescent cells in a group of five-month-old mice — when traits of ageing were evident. Another five months later, the mice showed increased muscle-fibre diameter and improved fitness performance compared with control mice. They also had enlarged fat deposits and decreased levels of senescence markers. So the removal of p16^{Ink4a}-positive cells also seems to be beneficial even later in life.

Baker and colleagues' elegant work is a proof-of-concept for the premise that therapeutic elimination of senescent cells may delay ageing-associated processes. Among the questions that therefore arise are whether these findings can be extrapolated to natural ageing, and whether the observed functional improvements are due to termination of further damage or damage reversal.

For therapeutic purposes, regulating the effects of senescent cells, rather than their abundance, may be more manageable. Of relevance to this is the large number of factors that are secreted by senescent cells. Some of these reinforce senescence in their cell of origin, whereas others affect nearby and even distant cells, disrupting vital processes such as tissue function and regeneration^{6,7} (Fig. 1). So identifying the factors that are secreted by senescent cells and contribute to ageing is a priority.

Deficiency of p16^{Ink4a} ameliorates traits of premature ageing, but does not completely prevent them or extend lifespan^{3,10}. Consistent with this, Baker *et al.* showed that features of ageing that are independent of p16^{Ink4a} remain unchanged³, supporting the notion that additional factors contribute to premature ageing. Investigations are therefore needed into whether such factors are directly involved in senescence, or instead are effector molecules of BubR1 function (such as the anaphase-promoting complex).

This paper³ may also provide a clue to why some organisms live longer than others. A provocative thought is that long-lived organisms may be equipped with an immune system that is better at eradicating senescent cells. Such communication between senescent cells and cells mediating immune surveillance is not unprecedented. In mouse liver, for example, injury to one cell type, hepatocytes, activates another cell type, hepatic stellate cells. The latter cells cause not only fibrosis but also senescence, thereby attracting the natural killer cells of the immune system to bring about

their own eventual demise and so limiting the damage¹¹. In humans, however, spontaneous clearance of cells that express p16^{Ink4a} abundantly is probably uncommon. As p16^{Ink4a} is often induced in precancerous lesions and even in tumours, eliminating p16^{Ink4a}-positive cells may come with benefits that go beyond ageing.

It was recently shown¹² that reactivation of the enzyme telomerase reverses degenerative traits. Baker *et al.* now find that clearing senescent cells delays age-related dysfunction in the setting of BubR1 insufficiency. Surely, researchers will now keenly explore these and other routes for targeting cellular senescence, in search of a viable strategy that would allow us to get older more healthily. ■

Daniel S. Peeper is at the Netherlands Cancer Institute, Division of Molecular Genetics,

1066 CX Amsterdam, the Netherlands.
e-mail: d.peeper@nki.nl

1. Kirkwood, T. B. L. & Austad, S. N. *Nature* **408**, 233–238 (2000).
2. Sharpless, N. E. & DePinho, R. A. *Nature Rev. Mol. Cell Biol.* **8**, 703–713 (2007).
3. Baker, D. J. *et al.* *Nature* **479**, 232–236 (2011).
4. Collado, M. & Serrano, M. *Nature Rev. Cancer* **10**, 51–57 (2010).
5. Kuilman, T., Michaloglou, C., Mooi, W. J. & Peeper, D. S. *Genes Dev.* **24**, 2463–2479 (2010).
6. Kuilman, T. & Peeper, D. S. *Nature Rev. Cancer* **9**, 81–94 (2009).
7. Rodier, F. & Campisi, J. J. *Cell Biol.* **192**, 547–556 (2011).
8. Campisi, J. *Cell* **120**, 513–522 (2005).
9. Pajvani, U. B. *et al.* *Nature Med.* **11**, 797–803 (2005).
10. Baker, D. J. *et al.* *Nature Cell Biol.* **10**, 825–836 (2008).
11. Krizhanovsky, V. *et al.* *Cell* **134**, 657–667 (2008).
12. Jaskelioff, M. *et al.* *Nature* **469**, 102–106 (2011).

NANOTECHNOLOGY

A molecular four-wheel drive

Nanoscale systems designed to imitate functions from the macroscopic world lead to a new appreciation of the complexity needed to actuate motion at the limits of miniaturization. A nanoscale 'car' is the latest example. SEE LETTER P.208

PAUL S. WEISS

A specially designed molecule that has four motorized 'wheels' is reported by Kudernac *et al.*¹ on page 208 of this issue. By depositing the molecules on a surface and providing them with sufficiently energetic electrons from the tip of a scanning tunnelling microscope, the authors were able to drive some of the molecules in a specific direction, much like a car with four-wheel drive. Previous examples of actuated single-molecule motion have been reported^{2–4}, but none with the complex action to continue moving in the same direction across a surface like the system devised, synthesized and operated by Kudernac and co-authors.

Biological systems have inspired the design, construction and operation of materials and nanometre-scale devices that have analogous functions on all scales⁴. In 'top-down' approaches, nanoscale components have been put together whose functions roughly mimic aspects of those of the kidneys, muscles, nose and so on^{4–6}. On the smallest scales, molecules have been designed and synthesized to demonstrate various capabilities, and, more importantly, to explore the limits of miniaturization, operation and efficiency for both natural and artificial machines. For example, single-molecule electronic switches and simple

molecules that move in response to various stimuli have been demonstrated^{2–4}. Typically, these synthetic systems are less flexible than their biological inspirations. This makes it easier to couple experimental results from such systems to theory and simulations ranging from quantum chemistry to mechanical engineering. From such studies, we ultimately hope to learn some of nature's secrets of robust operation, high energy-conversion efficiency and hierarchical function and action.

In such work, the forces that dominate the macroscopic world around us, such as gravity, are less important than the forces that rule the nanoscale and biological worlds, such as van der Waals and capillary forces. Nevertheless, there are similarities between our everyday world and the nanoscale world: a car must have not only the ability to accelerate, but also brakes and traction; likewise, nanoscale actuated motion depends on balancing applied forces with those that hold structures in place. Kudernac *et al.*¹ solve this problem, as others have, by balancing surface interactions and thermal energy with forces that can be applied through actuation, so that their molecules move only on stimulation, rather than by diffusion.

Scanning probe microscopes, especially scanning tunnelling microscopes (STMs), have been used to push, pull and place atoms,

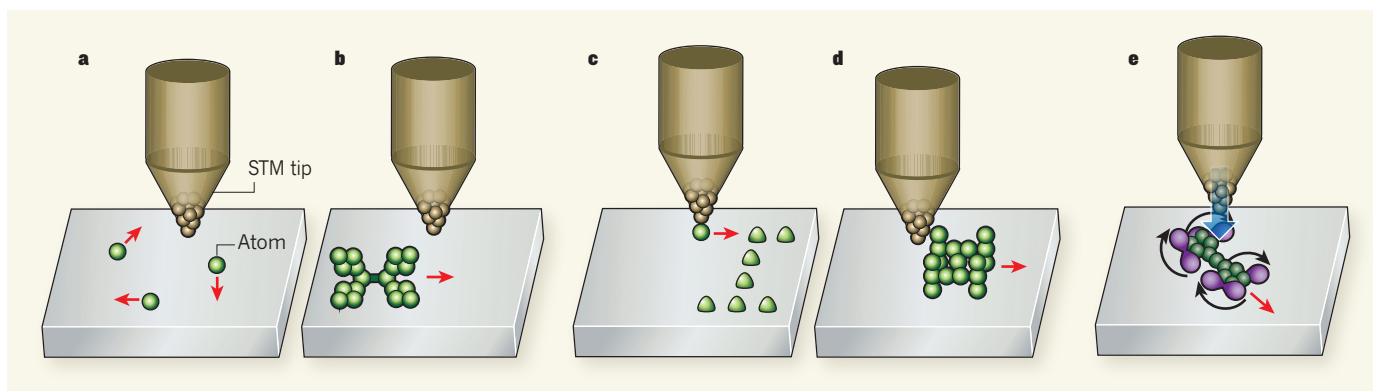


Figure 1 | Movement by microscope. Scanning tunnelling microscopes (STMs) have been pivotal to studies of atomic-scale motion. **a, b**, At first, STMs were used to follow the diffusion of atoms (**a**) and molecules (**b**) on surfaces. Red arrows indicate movement of atoms or molecules; the atomic structure of the end of the STM tip is shown. **c, d**, STMs were then used to

position atoms into precise structures (**c**) and to push molecules to study nanometre-scale interactions (**d**). **e**, Kudernac *et al.*¹ have now used an STM as an energy source for molecular motion — the STM provided electrons (blue arrow) to drive car-like molecules with four motorized ‘wheels’ (purple) across a surface. Black arrows indicate the turning of the wheels.

molecules and supramolecular assemblies on surfaces^{7–9} (Fig. 1). In complementary studies, mobile atoms and molecules on surfaces have been tracked using STMs to measure nanoscale interaction strengths between molecules, and between molecules and surfaces^{10–12}. In studies of both diffusion and manipulation, some of the molecules have been designed to include parts that test modes of motion. For example, do molecular wheels roll or slide? At least in some cases, they roll¹². And can synthetic molecules be made to run unidirectionally on tracks, like biomolecular motors? Thus far, they can do so only with carefully programmed tracks and fuel¹³.

Many of these molecules are analogous to familiar objects, and the similarities are often emphasized in schematic depictions of the molecules that leave out chemical groups and/or structural distortions caused by interactions of the molecules with surfaces. But such groups and distortions have significant consequences for their structures and motion on surfaces that often render our intuition about their behaviour incorrect. For example, although steps just one atom high on a surface can pose significant barriers for atoms, molecules and supramolecular structures moving on that surface, the relatively large molecules used in many studies distort so as to ‘climb’ up or down such steps with ease¹². In some cases, the presence of a moving molecule can even cause the atoms of the steps to reorganize, to make stronger contacts with molecules¹⁴.

With their inspired design of a four-wheeled molecule, Kudernac *et al.*¹ have gone a sizeable step further than previous studies. The wheels are actually molecular ratchets (rotary motors that turn in a single direction), some or all of which can be actuated to move when sufficiently energetic electrons are supplied by an STM tip. The authors observed that, on surfaces, the nanoscale car lurches forward by a fraction of a nanometre if each wheel rotates in the same direction. When driven repeatedly in this way, the molecules exhibited net forward

motion, although the route taken was not perfectly linear — presumably owing to mismatches between atoms of the surface and those of the wheels — and so the distance travelled was not quite as far as theoretically possible (see Fig. 2 of the paper¹). The authors used the STM not only to manipulate their molecules, but also, in a less perturbative mode, to image the progress of the molecules, an approach that has been used in earlier studies^{7–9,11}.

A complication of this work is that the chemical synthesis used by Kudernac *et al.* to make their molecules does not (yet) control the attachment of the wheels to ensure that they all turn in the same direction. In fact, only a fraction of the molecules synthesized had all four wheels optimally connected. Separating these cars from the others using chemical methods prior to deposition on a surface might be desirable, but this is prohibitively difficult at present. However, the properly constructed cars can be differentiated by function (that is, by their directional movement in response to stimulation) once they are deposited on a surface and probed individually.

In addition, the operation of the molecules depends on whether they land in the correct orientation on the surface — akin to having a car factory in which half the fully assembled vehicles are immobilized when they drop off the production line, because they land on their roofs or sides. The key difference is that, in chemical syntheses, trillions of molecules are synthesized in parallel. It has, however, been possible to design asymmetrical molecules in other systems that always orient themselves in a particular way upon deposition¹⁵. The orientational and other impediments in the present work¹ will, therefore, ultimately be cleared away through molecular-level understanding of the systems and further design constraints.

In developmental and evolutionary biology, ‘gains of function’ are key points in time. The work of Kudernac *et al.* represents just such a gain in function for nanotechnology:

molecules with multiple motors can now be synthesized, deposited and operated to move directionally across a surface. The authors’ system provides a tremendous opportunity to understand how molecular motors can interact, and how molecule–surface interactions affect molecular motion. It also, however, points out many of the complications that will come with creating increasingly complex syntheses and increasingly functional nanostructures. Nonetheless, we can expect further significant gains of function as our understanding of, and intuition for, the nanoscale world grows, and as accessible synthetic structures become ever more complex and sophisticated. ■

Paul S. Weiss is at the California NanoSystems Institute, and in the Departments of Chemistry & Biochemistry and Materials Science & Engineering, University of California, Los Angeles, Los Angeles, California 90095, USA. e-mail: psw@cnsi.ucla.edu

1. Kudernac, T. *et al.* *Nature* **479**, 208–211 (2011).
2. Donhauser, Z. J. *et al.* *Science* **292**, 2303–2307 (2001).
3. Kumar, A. S. *et al.* *Nano Lett.* **8**, 1644–1648 (2008).
4. Li, D. *et al.* *MRS Bull.* **34**, 671–681 (2009).
5. Nissenon, A. R., Ronco, C., Pergamit, G., Edelstein, M. & Watts, R. *Blood Purif.* **23**, 269–274 (2005).
6. Röck, F., Barsan, N. & Weimar, U. *Chem. Rev.* **108**, 705–725 (2008).
7. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524–526 (1990).
8. Otero, R., Rosei, F. & Besenbacher, F. *Annu. Rev. Phys. Chem.* **57**, 497–525 (2006).
9. Bartels, L., Meyer, G. & Rieder, K.-H. *Phys. Rev. Lett.* **79**, 697–700 (1997).
10. Linderoth, T. R., Horch, S., Lægsgaard, E., Stensgaard, I. & Besenbacher, F. *Phys. Rev. Lett.* **78**, 4978–4981 (1997).
11. Claridge, S. A., Schwartz, J. J. & Weiss, P. S. *ACS Nano* **5**, 693–729 (2011).
12. Shirai, Y., Osgood, A. J., Zhao, Y., Kelly, K. F. & Tour, J. M. *Nano Lett.* **5**, 2330–2334 (2005).
13. Sherman, W. B. & Seeman, N. C. *Nano Lett.* **4**, 1203–1207 (2004).
14. Rosei, F. *et al.* *Science* **296**, 328–331 (2002).
15. Ye, T., Takami, T., Wang, R., Jiang, J. & Weiss, P. S. *J. Am. Chem. Soc.* **128**, 10984–10985 (2006).

Distinct stem cells contribute to mammary gland development and maintenance

Alexandra Van Keymeulen^{1*}, Ana Sofia Rocha^{1*}, Marielle Ousset¹, Benjamin Beck¹, Gaëlle Bouvencourt¹, Jason Rock², Neha Sharma¹, Sophie Dekoninck¹ & Cédric Blanpain^{1,3}

The mammary epithelium is composed of several cell lineages including luminal, alveolar and myoepithelial cells. Transplantation studies have suggested that the mammary epithelium is maintained by the presence of multipotent mammary stem cells. To define the cellular hierarchy of the mammary gland during physiological conditions, we performed genetic lineage-tracing experiments and clonal analysis of the mouse mammary gland during development, adulthood and pregnancy. We found that in postnatal unperturbed mammary gland, both luminal and myoepithelial lineages contain long-lived unipotent stem cells that display extensive renewing capacities, as demonstrated by their ability to clonally expand during morphogenesis and adult life as well as undergo massive expansion during several cycles of pregnancy. The demonstration that the mammary gland contains different types of long-lived stem cells has profound implications for our understanding of mammary gland physiology and will be instrumental in unravelling the cells at the origin of breast cancers.

The mammary gland is composed of epithelial cells and mesenchymal cells, including adipocytes, fibroblasts, blood vessels and immune cells¹. Initially visible as placode-like structures, mammary glands are specified along the ventral epidermis during embryonic development and progressively invade the underlying mesenchyme, called the mammary fat pad. At puberty, the mammary gland expands considerably to form a highly branched tubular structure that progressively fills the fat pad. During pregnancy, the mammary gland expands further and the terminal end tubular structures differentiate into milk-producing cells. Two main cellular subtypes comprise the mammary gland epithelium: the basal myoepithelial cells and luminal cells, which can differentiate either into ductal cells or milk-producing cells (Supplementary Fig. 1). Whereas alveoli and luminal cells secrete the water and nutrients, the myoepithelial cells, through their contraction, guide the circulation of the milk throughout the ductal tree^{1–3}.

Different assays have been developed to define the differentiation potential of mammary epithelial cells (MECs)^{2,4,5}. *In vitro* assays indicated that both luminal cells and myoepithelial cells can be maintained with their lineage-restricted differentiation potential in a specific medium but only luminal cells can be forced to differentiate into myoepithelial cells upon medium switch⁶. Culture of fluorescence-activated cell sorting (FACS)-isolated human MECs gives rise to either luminal or myoepithelial colonies as well as some bipotent colonies^{6–8}. Culturing MECs as non-adherent cells, called mammospheres, allowed renewal and differentiation of cells with unipotent and bipotent differentiation potential⁹. Transplantation of primary MECs at limiting dilutions suggested the presence of multipotent mammary stem cells and more committed progenitors^{10,11}, and a single MEC is able to reconstitute an entire functional mammary gland in serial transplantation¹⁰. Transplantation of a single FACS-isolated MEC can reconstitute, although at low frequency, a normal mammary gland^{12,13}, indicating that rare multipotent mammary stem cells reside at the top of the cellular hierarchy within the mammary gland. Although

transplantation studies are important to define the differentiation potential of stem cells, these assays mimic a regenerative state that in certain circumstances forces stem cells to differentiate into lineages for which they usually do not contribute to under physiological conditions. For example, hair follicle bulge stem cells give rise to all epidermal lineages upon transplantation and wound healing, but only to hair follicle regeneration under physiological conditions¹⁴. The definitive demonstration that, under physiological conditions, multipotent stem cells are responsible for the development and adult maintenance of the mammary epithelium awaits genetic lineage-tracing experiments³. Here we developed novel lineage-tracing approaches in mice to decipher the cellular hierarchy of the mammary epithelium during physiological conditions.

Multipotent embryonic K14 progenitors

We first assessed the contribution of K14-derived cells to mammary gland development and adult life using K14-Cre/Rosa-YFP mice. The mammary placode arises from the embryonic epidermis at embryonic day 14 (E14). At E17, all MECs expressed K14 and were YFP⁺ in K14-Cre/Rosa-YFP mice and remained YFP⁺ thereafter (Supplementary Figs 2 and 3). During the early stage of mammary gland development, K14 expression encompassed both myoepithelial cells and a fraction of luminal cells (Fig. 1a–c). At birth, the mammary gland consisted of a tubular epithelial structure composed of basal myoepithelial cells expressing K5, K14 and SMA and luminal cells expressing K8 and K19. At the beginning of puberty and thereafter, K14 expression was restricted to the myoepithelial lineage (Supplementary Figs 1 and 2). FACS analysis of mammary gland from K14-Cre/Rosa-YFP mice revealed that all YFP-labelled cells expressed CD24, as previously suggested¹⁵, and could be divided into two populations: CD29^{Hi}CD24⁺ enriched for myoepithelial cells and CD29^{Lo}CD24⁺ enriched for luminal cells^{12,13} (Supplementary Figs 4 and 5). Because K14-Cre is expressed in myoepithelial cells after birth, we investigated whether all MECs derive from embryonic K14⁺ progenitors by administering

¹Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium. ²Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. ³Welbio, Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium.

*These authors contributed equally to this work.

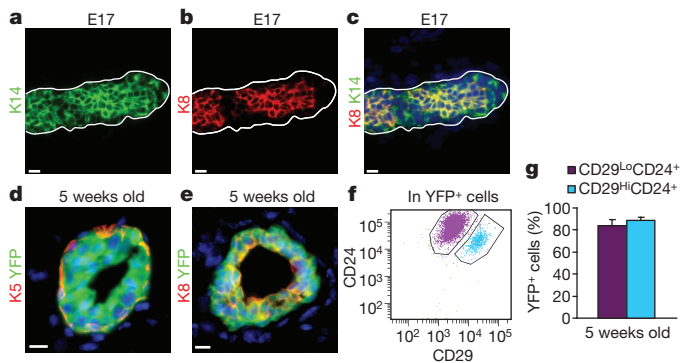


Figure 1 | All mammary epithelial lineages derive from embryonic K14-expressing progenitors. **a–c**, Immunostaining of K8 and K14 in mammary glands at E17. **d, e**, Immunostaining of K5 (**d**) or K8 (**e**) and YFP in K14-rtTA/TetO-Cre/Rosa-YFP MECs 5 weeks after doxycycline administration. **f**, FACS analysis of CD24 and CD29 expression in Lin[−]YFP⁺ cells 5 weeks after doxycycline administration. **g**, YFP expression in Lin[−]CD29^{Lo}CD24⁺ and Lin[−]CD29^{Hi}CD24⁺ populations 5 weeks after doxycycline administration, showing that embryonic K14 tracing marked the vast majority of MECs ($n = 3$ mice). Scale bars, 10 µm; error bars indicate s.e.m.

a single dose of doxycycline to pregnant K14-rtTA/TetO-Cre/Rosa-YFP mice at E17 to label MECs during development (Supplementary Fig. 6). Analysis of the mammary gland at puberty showed that embryonic labelling marked the majority of MECs, including both myoepithelial cells and luminal cells, demonstrating that embryonic K14⁺ progenitors give rise to all MEC lineages (Fig. 1d–g).

K14 myoepithelial stem cells

The transcriptional profiling of CD29^{Hi}CD24⁺ cells revealed that the putative multipotent mammary stem cells^{12,13} are enriched for K5 and K14 (refs 12, 13, 16, 17). To determine whether postnatal K14⁺ cells contain multipotent mammary stem cells, we performed inducible genetic lineage-tracing experiments of K14-expressing cells during puberty and in adult virgin mice. Doxycycline administration over 5 days in K14-rtTA/TetO-Cre/Rosa-YFP mice induced YFP expression in about 40% of myoepithelial cells but did not label luminal cells (Fig. 2a, b, e and Supplementary Figs 7 and 8). Surprisingly, 10 weeks after doxycycline administration, YFP was still exclusively expressed by myoepithelial cells (Fig. 2c–e and Supplementary Figs 7 and 8). To rule out that the remaining CD29^{Hi}CD24⁺ myoepithelial cell population that was not labelled in the previous experiments contains multipotent mammary stem cells, we administered doxycycline continuously to K14-rtTA/TetO-Cre/Rosa-YFP mice during the whole process of pubertal development and found that almost all myoepithelial cells (>97%) but no luminal cells were labelled (Fig. 2f, g and Supplementary Fig. 7m). These data demonstrate that the K14-expressing cells do not contribute to the luminal lineage during mammary gland expansion that occurred during pubertal development.

Administration of a low dose of doxycycline to K14-rtTA/TetO-Cre/Rosa-YFP 4-week-old mice resulted in the labelling of isolated myoepithelial cells 1 week after doxycycline administration. These cells were maintained for several weeks and about 10% of them expanded over time (Fig. 2h, i and Supplementary Fig. 9). YFP⁺ myoepithelial cells expanded further during pregnancy and lactation. Some of them escaped mammary gland involution and reinitiated another cycle of expansion during the following pregnancy and were still present after the second mammary gland involution (Fig. 2j–l and Supplementary Fig. 10). The proportion of YFP-labelled cells was stable over time (Fig. 2l), showing that these unipotent stem cells undergo long-term self-renewal and are not replaced by multipotent stem cells over time.

Doxycycline administration to K14-rtTA/TetO-Cre/Rosa-YFP mice at postnatal day 1 (P1) marked mostly myoepithelial cells and

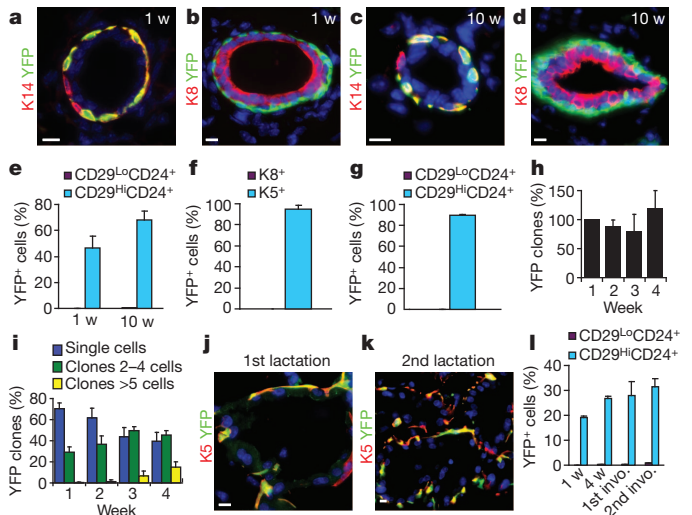


Figure 2 | K14-expressing stem cells ensure mammary myoepithelial lineage expansion and maintenance during puberty, adult life and pregnancy. **a–d**, Immunostaining of K14 (**a, c**) or K8 (**b, d**) and YFP 1 week (**a, b**) and 10 weeks (**c, d**) after doxycycline administration. **e**, Percentage of YFP⁺ cells in Lin[−]CD29^{Hi}CD24⁺ and Lin[−]CD29^{Lo}CD24⁺ populations 1 week and 10 weeks after doxycycline administration at puberty ($n = 3$ mice per time point). **f, g**, YFP expression in myoepithelial cells and luminal cells as determined by immunofluorescence (**f**) and FACS analysis (**g**) after 1 month of continuous doxycycline administration during puberty ($n = 2$ mice per time point). **h**, Percentage of YFP myoepithelial clones observed at different time points. **i**, Temporal distribution of clone size observed as single cell, 2–4 cells and 5 cells or more ($n =$ at least 470 clones from 3 different mice were counted at each time point). **j, k**, Immunostaining of K5 and YFP in MECs during first and second lactation after 1 mg doxycycline administration at puberty. **l**, Percentage of YFP⁺ cells in MECs at different time points after doxycycline administration at puberty ($n =$ at least 2 mice per time point). Scale bars, 10 µm unless stated; error bars indicate s.e.m.

only a small proportion of luminal cells (Supplementary Fig. 11), indicating that the lineage restriction of K14-expressing cells occurred rapidly after birth.

To determine whether K14-rtTA/TetO-Cre targeted a subpopulation of myoepithelial cells distinct from the putative mammary stem cells, we repeated these experiments using K5-CreER mice, another marker preferentially expressed by putative multipotent mammary stem cells^{12,13,16,17}. Tamoxifen administration to P1 and 4-week-old K5-CreER/Rosa-YFP mice initially labelled exclusively myoepithelial cells, giving rise several weeks later to myoepithelial cells only and not luminal cells. Similarly to K14-rtTA/TetO-Cre targeted cells, a fraction of K5-CreER targeted cells was able to clonally expand during puberty and pregnancy (Supplementary Figs 12–15), confirming the presence of myoepithelial stem cells and the apparent absence of multipotent mammary stem cells among postnatal MECs expressing K5 and K14.

Lgr5, a marker of epithelial stem cells in different tissues¹⁸, has been reported to be enriched in the CD29^{Hi}CD24⁺ population¹². Lgr5 is expressed in only 2–3% of MECs, localized to the nipple region. At puberty, the vast majority of Lgr5⁺ cells within the mammary gland were myoepithelial cells, although a small fraction of Lgr5⁺ cells was also luminal (Supplementary Fig. 16a–d). Tamoxifen administration to 4-week-old Lgr5-GFP-CreER/Rosa-Tomato mice preferentially labelled myoepithelial cells, giving rise to myoepithelial cells several weeks later (Supplementary Fig. 16e–m). Functionally, Lgr5⁺ myoepithelial cells were indistinguishable from K14⁺ cells except for their clustering close to the nipple region. The only difference between Lgr5 and K14 is the Lgr5 expression and tracing of rare luminal cells.

K8 luminal stem cells

Lineage-tracing experiments using WAP-Cre mice, which is expressed and active in luminal cells during pregnancy, identified

long-lived cells that display the ability to clonally expand during pregnancy and give rise to luminal and alveolar cells, and are therefore called parity-induced cells¹⁹. *Ex vivo* culture of WAP-Cre mammary explants, in the presence of several growth factors, suggested that cells with similar renewal and differentiation potential as parity-induced mammary progenitors may already exist in nulliparous mice²⁰. To determine whether, under physiological conditions without *ex vivo* manipulations, luminal cells contain stem cells before pregnancy and whether these cells represent unipotent or multipotent stem cells, as has been previously suggested for human MECs⁶, we generated transgenic mice expressing CreER in the luminal lineage using the K8 promoter (Supplementary Fig. 17a). Administration of tamoxifen to K8-CreER/Rosa-YFP mice at P1 only traced luminal cells (Supplementary Fig. 17). Similarly, tamoxifen administration to 4-week-old and adult virgin mice induced YFP expression only in luminal cells (Fig. 3a, b and Supplementary Figs 18 and 19), and after 10 weeks, YFP⁺ cells had expanded but were still luminal cells (Fig. 3c, d and Supplementary Figs 18 and 19), indicating that K8⁺ cells contain luminal stem cells.

Clonal analysis of K8⁺ cells by administration of a low dose of tamoxifen revealed that not all YFP⁺ cells persist long-term and at 4 weeks after tamoxifen administration 40% of YFP luminal clones were lost, indicating that K8-CreER also targets more committed luminal cells (Fig. 3e–h and Supplementary Figs 20–22). Temporal analysis of clone size revealed that about 10% of the YFP clones contained more than 5 YFP⁺ cells, some of which could be even much larger (Fig. 3g and Supplementary Fig. 21), consistent with the targeting of luminal stem cells that clonally expand and participate in luminal cell expansion during puberty and maintenance during adult life.

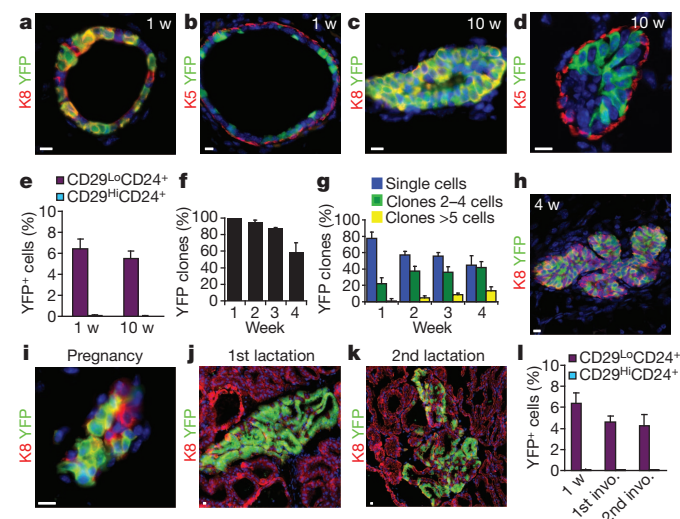


Figure 3 | K8-expressing stem cells ensure mammary luminal lineage expansion and maintenance during puberty, adult life and pregnancy. a–d, Immunostaining of K8 (a, c) or K5 (b, d) and YFP in MECs 1 week (a, b) and 10 weeks (c, d) after 15 mg tamoxifen administration at puberty. e, Percentage of YFP⁺ cells in Lin[−]CD29^{Lo}CD24⁺ and Lin[−]CD29^{Hi}CD24⁺ populations 1 week and 10 weeks after 1 mg tamoxifen administration to K8-CreER/Rosa-YFP mice at puberty ($n = 7$ and 4 mice at 1 week and 10 weeks). f, g, Number of YFP clones (f) and distribution of clone sizes (g) observed as single cell, 2–4 cells, and 5 or more cells at different time points after 1 mg tamoxifen administration at puberty ($n =$ at least 526 clones analysed from three different mice at each time point). h–k, Immunostaining of K8 (red) and YFP (green) in the mammary gland 4 weeks after induction (h), during pregnancy (i), lactation (j) and second lactation (k) in K8-CreER/Rosa-YFP mice treated with 1 mg tamoxifen at puberty. l, Percentage of YFP⁺ cells in MECs at different time points after tamoxifen administration in K8-CreER/Rosa-YFP mice treated with 1 mg tamoxifen at puberty ($n =$ at least 2 mice per time point). Scale bars, 10 μ m unless stated; error bars indicate s.e.m.

To establish further the renewal capacities of K8⁺ luminal stem cells and their contribution to milk-producing cells, we induced clonal YFP expression in luminal cells during puberty and followed their fate during pregnancy and lactation (Fig. 3i–l and Supplementary Fig. 23). During pregnancy, only clones of YFP⁺ luminal cells were found. During lactation, very large YFP clones were observed with some lobules that were almost entirely YFP⁺ whereas others were either negative or mosaic for YFP expression (Fig. 3j and Supplementary Fig. 23). K8-CreER-targeted cells differentiated into both luminal and milk-producing cells, as revealed by the co-expression of YFP with K8 and NaPiIIB in the fat-milk-producing cells; no myoepithelial cells were YFP⁺ (Fig. 3j and Supplementary Fig. 23d). After involution, some YFP⁺ cells persisted and were able to reinitiate another round of expansion during the following pregnancy and lactation and to escape cell death during the second involution (Fig. 3k and Supplementary Fig. 23). Even after three consecutive cycles of pregnancy and lactation, K8-derived cells were found in luminal cells and milk-producing cells (Supplementary Fig. 23j–m). The percentage of YFP⁺ cells was stable over time (Fig. 3l), indicating that these cells are self-renewing long-term and are not progressively replaced by multipotent stem cells.

K18 committed luminal cells

Different studies suggested that luminal cells are composed of morphologically distinct cell types that are thought to display different proliferation and differentiation capacities^{21–23}. We used another luminal inducible CreER (K18-CreER)²⁴ to determine whether all luminal cells presented similar renewal and differentiation potential as compared to K8-CreER-targeted cells (Supplementary Figs 24–28). Administration of tamoxifen to 4- and 8-week-old K18-CreER/Rosa-YFP mice resulted in a patchy expression of YFP in the mammary gland. Only luminal cells were initially labelled and 10 weeks after tamoxifen administration, YFP-marked luminal cells were still present whereas no myoepithelial cells were YFP⁺ (Supplementary Figs 24 and 25). Clonal analysis during puberty and in adult virgin mice revealed no sign of important clonal expansion, even during pregnancy and lactation (Supplementary Figs 26–28). Together, these data suggest that the luminal cells targeted by K18-CreER display a low cellular turnover, and should be considered as more committed luminal cells, possibly representing the cells that are lost and/or failed to expand in K8 lineage-tracing experiments.

Transplantation and stem cell differentiation

To clarify the discrepancy between the results obtained in transplantation assays^{12,13} and our lineage-tracing experiments, we performed mammary reconstitution assays with cells labelled by our different myoepithelial- and luminal-specific Cre. Transplantation of 10^6 dissociated mammary gland cells—which represent about 5×10^4 living MECs from 4-week-old K14-Cre/Rosa-YFP mice—into the fat pad of NOD/SCID mice allowed the reconstitution of a morphologically normal mammary gland with YFP⁺ myoepithelial cells and luminal cells (Supplementary Fig. 29a–d).

To determine whether the transplantation procedure itself can promote the differentiation potential of YFP-labelled stem cells into the other lineage, we induced YFP expression in myoepithelial cells by administering doxycycline to 4-week-old K14-rtTA/TetO-Cre/Rosa-YFP mice. One week later the mammary gland was dissociated into single cells and a mixture of YFP⁺ myoepithelial cells together with unlabelled luminal cells was transplanted into the mammary fat pad of NOD/SCID mice (Supplementary Fig. 29a). Seven weeks after transplantation, the grafted cells regenerated a new mammary gland expressing YFP in the ducts and growing alveoli (Fig. 4a). Microscopic examination revealed that the vast majority of YFP⁺ cells were myoepithelial cells (Fig. 4b, c and Supplementary Fig. 29e–h) and only very rare clones expressing YFP in both myoepithelial cells and luminal cells were identified (Supplementary Fig. 29i, j and

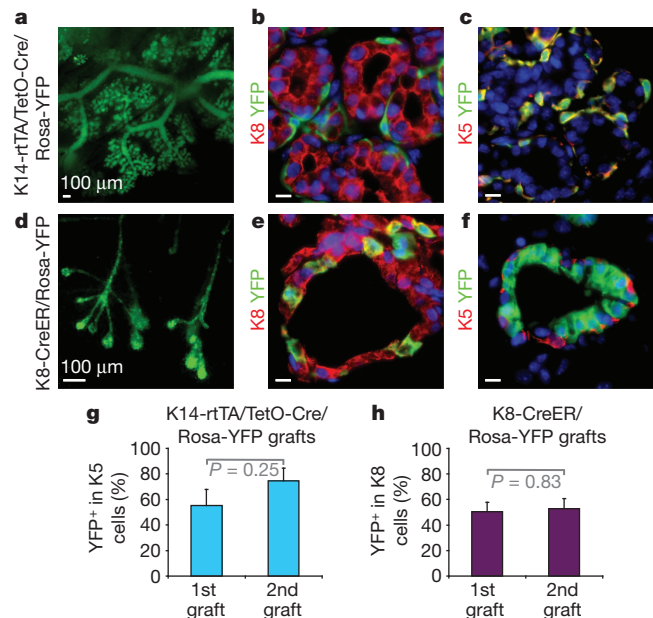


Figure 4 | Myoepithelial and luminal stem cells maintain their unipotent fate in mammary reconstitution assay when co-transplanted in non-limiting conditions. **a**, Whole-mount of the reconstituted mammary gland after transplantation of mammary cells in which myoepithelial cells have been induced to express YFP. **b, c**, Immunostaining of K8 (**b**) or K5 (**c**) and YFP in the K14-rtTA/TetO-Cre/Rosa-YFP graft. **d**, Whole-mount of the reconstituted mammary gland after the transplantation of breast cell suspensions in which a fraction of luminal cells has been induced to express YFP. **e, f**, Immunostaining of K8 (**e**) or K5 (**f**) and YFP in the K8-CreER/Rosa-YFP graft. **g, h**, Percentage of YFP⁺ cells in K5⁺ (**g**) and K8⁺ (**h**) cells in the first and second transplants quantified by immunofluorescence in serial transplantation of unsorted cells from K14-rtTA/TetO-Cre/Rosa-YFP (**g**) and K8-CreER/Rosa-YFP (**h**) mice. Scale bars, 10 μ m unless stated; error bars indicate s.e.m.

Supplementary Table 1). Similar results were obtained after the transplantation of MECs labelled by K5-CreER (Supplementary Fig. 29k–m). Transplantation of YFP-marked luminal cells together with unlabelled myoepithelial cells after tamoxifen administration to K8-CreER/Rosa-YFP mice resulted in the regeneration of a new mammary gland expressing YFP in the duct and growing alveoli, but in this case only luminal cells and not myoepithelial cells were YFP⁺ (Fig. 4d–f). Both luminal and myoepithelial YFP-marked cells could be serially transplanted without a significant decrease in the overall YFP chimaerism and with similar differentiation potential as obtained in the first transplantation. This suggests that both types of stem cells possess a sustained renewal potential and are not progressively replaced by multipotent stem cells (Fig. 4g, h and Supplementary Fig. 30). These experiments clearly demonstrate that luminal and myoepithelial stem cells actively participate in epithelial regeneration in mammary reconstitution assays while still exhibiting the similar lineage-restricted differentiation as observed in unperturbed mammary gland when the two types of unipotent stem cells are present together at non-limiting dilutions.

To determine whether the presence of only one type of mammary stem cell during the transplantation procedure can expand their differentiation potential, FACS-isolated YFP⁺ myoepithelial or luminal cells were transplanted into the mammary fat pad of NOD/SCID mice (Supplementary Fig. 31a). As previously shown^{12,13}, transplantation of FACS-isolated CD29^{Lo}CD24⁺ luminal cells alone failed to reconstitute a new mammary gland upon transplantation whereas CD29^{Hi}CD24⁺ myoepithelial cells alone were able to regenerate a new mammary gland independently of the presence of luminal cells (Fig. 5a–c and Supplementary Table 1). Finally, we assessed whether there is a critical threshold of luminal/myoepithelial cell ratio that

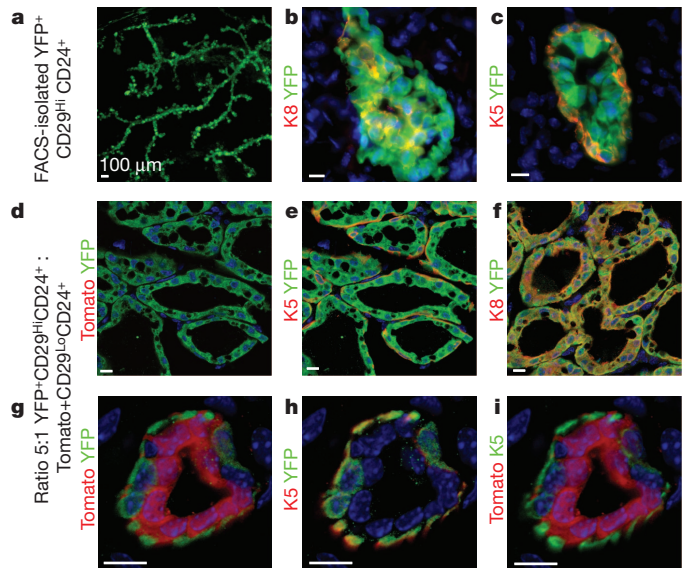


Figure 5 | Myoepithelial but not luminal stem cells can be forced to adopt a multipotent fate in mammary reconstitution assays. **a**, Whole-mount of the graft obtained after the transplantation of FACS-isolated YFP⁺CD29^{Hi}CD24⁺ cells. **b, c**, Immunostaining of K8 (**b**) and K5 (**c**) and YFP. **d–i**, Transplantation of 10⁴ FACS-isolated YFP⁺CD29^{Hi}CD24⁺ cells together with 2,000 Tomato⁺CD29^{Lo}CD24⁺ cells into the mammary fat pads of NOD/SCID mice. Immunostaining of Tomato and YFP (**d, g**), K5 and YFP (**e, h**), K8 and YFP (**f**) and K5 and Tomato (**i**). Scale bars, 10 μ m unless stated.

uncovers the bipotentiality of stem cells. Transplantation of FACS-isolated YFP⁺ myoepithelial cells and Tomato⁺ luminal cells showed that decreasing the luminal/myoepithelial cell ratio to 1/5, which is about ten times lower than physiological conditions, markedly increased the ability of myoepithelial cells to differentiate into luminal cells, as about half of the grafts were exclusively derived from YFP⁺ myoepithelial cells (Fig. 5d–f and Supplementary Table 1), whereas the other half was composed of a mixture of YFP⁺ myoepithelial cells and Tomato⁺ luminal cells (Fig. 5g–i and Supplementary Fig. 31).

Discussion

Our study shows that the mammary gland initially develops from multipotent embryonic K14⁺ progenitors, which give rise to both myoepithelial cells and luminal cells, as has been suggested previously²⁵. During puberty and homeostasis, the expansion and maintenance of each lineage is ensured by the presence of two types of lineage-restricted stem cell, able to differentiate into either myoepithelial or luminal lineages, rather than being maintained by rare multipotent stem cells (Supplementary Fig. 32). Our data cannot rule out that some rare multipotent stem cells that were not targeted by the multiple different Cre lines used in this study exist in the mammary epithelium. However, the long-term maintenance and stable frequency of YFP⁺ cells, their ability to be serially transplanted as well as their massive expansion during pregnancy and lactation clearly show that unipotent stem cells are not replaced by multipotent stem cells over time and that the contribution of multipotent stem cells to mammary gland morphogenesis and adult maintenance, if it exists, is very limited under physiological conditions.

Consistent with our data, previous experiments in which labelled and non-labelled MECs have been transplanted into the fat pad showed that the new mammary glands were composed of mixture of labelled and unlabelled cells^{11,19,26–29}, suggesting that multiple progenitors contribute to mammary gland development and adult maintenance rather than being performed by rare stem cells. In humans, analysis of X-chromosome inactivation indicates that breast epithelium is organized into multiple discrete regions sharing the same inactive X chromosome

and suggested the existence of multiple stem cells scattered throughout the gland³⁰.

Our study also demonstrates that transplantation assays, although extremely informative about the differentiation potential of tissue-specific stem cells, can be misleading in extrapolating the differentiation potential of stem cells under physiological conditions. These results may explain why genetically altered strains of mice that lack basal mammary stem cells in transplantation assays show no particular defects in development or pregnancy^{31,32}. The very low frequency of luminal differentiation of transplanted unipotent myoepithelial stem cells when luminal cells are present within the graft at the physiological ratio, and the increase in the multipotent differentiation of myoepithelial cells when luminal cells were depleted, suggested that either luminal cells restrict the differentiation potential of myoepithelial stem cells or that the differentiation of myoepithelial stem cells into luminal cells is a relatively rare event compared to the natural differentiation of luminal stem cells, and consequently these rare multipotent clones were out-competed by clones originating from luminal stem cells in these non-limited dilution experiments. The reason why myoepithelial cells can adopt a multipotent fate and are able to regenerate a complete mammary gland upon transplantation remains elusive. One possibility would be that mammary reconstitution assays recapitulate the process of mammary gland development and adult K14/K5⁺ stem cells are more prone to dedifferentiate into K14/K5 embryonic multipotent mammary progenitors under these conditions.

METHODS SUMMARY

YFP expression was induced in K8-CreER/Rosa-YFP, K18-CreER/Rosa-YFP and K5-CreER/Rosa-YFP female mice by intraperitoneal tamoxifen injection and in K14-rtTA/TetO-Cre/Rosa-YFP mice by oral administration of doxycycline food or by intraperitoneal injection. Immunostaining was performed as described²⁴. Mammary glands were dissected and lymph nodes removed. Tissues were cut in pieces of 1 mm³ and digested in HBSS plus 300 U ml⁻¹ collagenase plus 300 µg ml⁻¹ hyaluronidase for 2 h at 37 °C under shaking. EDTA was added for 10 min, followed by trypsin-EGTA for 2 min. Cell labelling, flow cytometry and sorting were performed as described^{12,13}. Dead cells were excluded with DAPI; CD45⁺, CD31⁺ and CD140a⁺ cells were excluded (Lin⁻) before analysis. Unsorted mammary cells or FACS-isolated cells were re-suspended in 10 µl PBS and injected into the cleared mammary fat pad of NOD/SCID female mice. Recipient mice were mated 4 weeks after the transplantation and analysed 3 weeks later.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 March; accepted 21 September 2011.

Published online 9 October 2011.

- Watson, C. J. & Khaled, W. T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, 995–1003 (2008).
- Smalley, M. & Ashworth, A. Stem cells and breast cancer: A field in transit. *Nature Rev. Cancer* **3**, 832–844 (2003).
- Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* **23**, 2563–2577 (2009).
- Visvader, J. E. & Smith, G. H. Murine mammary epithelial stem cells: Discovery, function, and current status. *Cold Spring Harb. Perspect. Biol.* **3**, 1–14 (2011).
- Stingl, J. Detection and analysis of mammary gland stem cells. *J. Pathol.* **217**, 229–241 (2009).
- Pechoux, C., Gudjonsson, T., Ronnov-Jessen, L., Bissell, M. J. & Petersen, O. W. Human mammary luminal epithelial cells contain progenitors to myoepithelial cells. *Dev. Biol.* **206**, 88–99 (1999).
- Stingl, J., Eaves, C. J., Kuusk, U. & Emsman, J. T. Phenotypic and functional characterization *in vitro* of a multipotent epithelial cell present in the normal adult human breast. *Differentiation* **63**, 201–213 (1998).
- Gudjonsson, T. *et al.* Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev.* **16**, 693–706 (2002).
- Dontu, G. *et al.* *In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270 (2003).
- Kordon, E. C. & Smith, G. H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921–1930 (1998).
- Smith, G. H. Experimental mammary epithelial morphogenesis in an *in vivo* model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res. Treat.* **39**, 21–31 (1996).

- Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
- Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
- Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature Rev. Mol. Cell Biol.* **10**, 207–217 (2009).
- Sleeman, K. E., Kendrick, H., Ashworth, A., Isacke, C. M. & Smalley, M. J. CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. *Breast Cancer Res.* **8**, R7 (2006).
- Lim, E. *et al.* Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21 (2010).
- Kendrick, H. *et al.* Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* **9**, 591 (2008).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- Wagner, K. U. *et al.* An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* **129**, 1377–1386 (2002).
- Booth, B. W., Boulanger, C. A. & Smith, G. H. Alveolar progenitor cells develop in mouse mammary glands independent of pregnancy and lactation. *J. Cell. Physiol.* **212**, 729–736 (2007).
- Chepko, G. & Smith, G. H. Three division-competent, structurally-distinct cell populations contribute to murine mammary epithelial renewal. *Tissue Cell* **29**, 239–253 (1997).
- Rudland, P. S. & Hughes, C. M. Immunocytochemical identification of cell types in human mammary gland: variations in cellular markers are dependent on glandular topography and differentiation. *J. Histochem. Cytochem.* **37**, 1087–1100 (1989).
- Fernandez-Gonzalez, R. *et al.* Mapping mammary gland architecture using multi-scale *in situ* analysis. *Integr. Biol.* **1**, 80–89 (2009).
- Van Keymeulen, A. *et al.* Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis. *J. Cell Biol.* **187**, 91–100 (2009).
- Liu, X. *et al.* Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proc. Natl Acad. Sci. USA* **104**, 12111–12116 (2007).
- Srinivasan, K., Strickland, P., Valdes, A., Shin, G. C. & Hinck, L. Netrin-1/neogenin interaction stabilizes multipotent progenitor cap cells during mammary gland morphogenesis. *Dev. Cell* **4**, 371–382 (2003).
- Naylor, S. *et al.* Retroviral expression of Wnt-1 and Wnt-7b produces different effects in mouse mammary epithelium. *J. Cell Sci.* **113**, 2129–2138 (2000).
- Welm, B. E. *et al.* Sca-1^{pos} cells in the mouse mammary gland represent an enriched progenitor cell population. *Dev. Biol.* **245**, 42–56 (2002).
- Welm, B. E., Dijkgraaf, G. J., Bledau, A. S., Welm, A. L. & Werb, Z. Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. *Cell Stem Cell* **2**, 90–102 (2008).
- Tsai, Y. C. *et al.* Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis. *Cancer Res.* **56**, 402–404 (1996).
- Badders, N. M. *et al.* The Wnt receptor, Lrp5, is expressed by mouse mammary stem cells and is required to maintain the basal lineage. *PLoS ONE* **4**, e6594 (2009).
- Taddei, I. *et al.* β 1 integrin deletion from the basal compartment of the mammary epithelium affects stem cells. *Nature Cell Biol.* **10**, 716–722 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank our colleagues who provided us with reagents, which are cited in the text, and B. Hogan for sharing unpublished mice. We thank our colleagues from the Blanpain laboratory and C. Govaerts for their comments on the manuscript. We thank J. Rosen for discussion and M. Van Lohuizen and K. Nacerddine for their help with the transplantation assay. We thank F. Bollet-Quivogne and J.-M. Vanderwinden for their help with confocal imaging. C.B. and A.V.K. are chercheur qualifié, B.B. is chargé de recherche and M.O. is a collaborateur scientifique of the FRF/FNRS. A.S.R. is supported by TELEVE and the Portuguese Science Foundation (FCT). N.S. is supported by the Fondation Contre le Cancer. J.R. is supported by the grant F32HL102920. C.B. is an investigator of Welbio. This work was supported by the FNRS, TELEVE, the program d'excellence CIBLES of the Wallonia Region, a research grant from the Fondation Contre le Cancer, the ULB fondation, the fond Gaston Ithier, a starting grant of the European Research Council (ERC) and the EMBO Young Investigator Program.

Author Contributions C.B., A.V.K., A.S.R. designed the experiments and performed data analysis. A.V.K., A.S.R. and M.O. performed most of the experiments, J.R. generated the K5-CreER knockin mice, B.B., S.D. and A.V.K. performed the FACS analysis and cell sorting. G.B. and N.S. provided technical support. C.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B. (Cedric.Blanpain@ulb.ac.be).

METHODS

Mice. Rosa-YFP³³, Rosa-Tomato³⁴, and Lgr5-GFP-CreER mice¹⁸ were obtained from the Jackson Laboratory. K14-Cre transgenic mice³⁵ and K14-rtTA³⁶ mice were provided by E. Fuchs. TetO-Cre mice³⁷ were provided by A. Nagy. The generation of K18-CreER mice was as previously described²⁴. Mice colonies were maintained in a certified animal facility in accordance with European guidelines. These experiments were approved by the local ethical committee (CEBEA).

Generation of K8-CreER mice. The CreERT2 fragment (supplied by P. Chambon) preceded by the β -globin intron and followed by a SV40 polyA signal was subcloned into pBluescript II SK⁺. The 3.5-kb sequence upstream the ATG codon of the murine K8 gene, obtained from the BAC clone RP23-254K21 (BACPAC Resources Center, Children's Hospital Oakland Research Institute) using the forward primer 5'-GGTGGATCACTTGCCCCCTCCGTTTG-3' and the reverse primer 5'-GGGACAGCGCCAGCGAAGGCC-3', was cloned upstream of the β -globin intron. The resulting K8-CreER fragment of 6.3 kb was released from the backbone by NotI digestion and was microinjected into fertilized oocytes to generate transgenic mice (Jacquemin laboratory). Seven transgenic founders were first identified by PCR, out of 27 mice born. Expression profiles of the K8-CreER founders were screened with reporter Rosa-YFP mice. Four founders expressed the YFP in cells expressing the endogenous K8, and one founder K8 was used throughout this study.

Generation of K5-CreER mice. The CreERT2 fragment, preceded by IRES, was inserted into the 3' UTR of *Krt5* in 129-derived ES cells. Correctly targeted cells were injected into B6 blastocysts. The *neo* selection cassette was removed by crossing a chimaeric K5-CreER male to a β -actin-Flp female (B6;SJL-Tg(ACTFLPe)9205Dym/J). First generation of animals backcrossed to B6 were crossed with Rosa-YFP mice.

Targeting YFP or Tomato expression. K14-Cre/Rosa-YFP female mice express YFP in all cells derived from K14-expressing cells, whereas K14-Cre/Rosa-Tomato mice express Tomato in all cells derived from K14-expressing cells. For lineage tracing induced at 4 weeks or at 8 weeks, K8-CreER/Rosa-YFP, K18-CreER/Rosa-YFP, K5-CreER/Rosa-YFP and Lgr5-GFP-CreER/Rosa-Tomato female mice were induced with 15 mg of tamoxifen (Sigma, diluted in sunflower seed oil, Sigma) by intraperitoneal injection (three injections of 5 mg every other day) and K14-rtTA/TetO-Cre/Rosa-YFP female mice were induced by oral administration of doxycycline food diet (1 g kg⁻¹, BIO-SERV) during 5 days, or 1 month. For lineage tracing induced at E17, pregnant females of K14-rtTA/TetO-Cre/Rosa-YFP mice were fed with a doxycycline food diet at E17 for 3 days. For lineage tracing induced at birth, K8-CreER/Rosa-YFP and K5-CreER/Rosa-YFP newborn mice were induced with 125 μ g tamoxifen (5 μ l of 25 mg ml⁻¹ solution) and K14-rtTA/TetO-Cre/Rosa-YFP female mice were induced with 25 μ g doxycycline (Sigma) (5 μ l of 5 mg ml⁻¹ solution diluted in sterile PBS) by intraperitoneal injection. For two-dimensional clonal analysis, K8-CreER/Rosa-YFP and K18-CreER/Rosa-YFP female mice were induced respectively with 1 mg and 10 mg tamoxifen and K14-rtTA/TetO-Cre/Rosa-YFP female mice were induced with 1 mg of doxycycline by intraperitoneal injection. For clonal analysis by confocal microscopy, K8-CreER/Rosa-YFP and K5-CreER/Rosa-YFP female mice were induced respectively with 0.2 and 7.5 mg tamoxifen by intraperitoneal injection.

Histology and immunostaining. Dissected mammary glands were pre-fixed for 2 h in 4% paraformaldehyde at room temperature. Tissues were washed three times with PBS for 5 min and incubated overnight in 30% sucrose in PBS at 4 °C. Tissues were then embedded in OCT and kept at -80 °C. Sections of 5 μ m (for epifluorescence analysis) or 50 μ m (for confocal analysis) were cut using a HM560 Microm cryostat (Mikron Instruments).

Five-micrometre sections were incubated in blocking buffer (5% NDS/1% BSA/0.2% Triton in PBS) for 1 h at room temperature. The different primary antibody combinations were incubated overnight at 4 °C. Sections were then rinsed three times for 5 min in PBS and incubated with proper secondary antibodies diluted at 1:400 in blocking buffer for 1 h at room temperature. Fifty-micrometre sections were incubated in blocking buffer (5% NDS/1% BSA/0.5% Triton/0.1% Tween 20 in PBS) for 2 h at room temperature. The different primary antibody combinations were incubated overnight at 4 °C. Sections were then rinsed three times in PBS for 45 min and incubated with proper secondary antibodies diluted at 1:400 in blocking buffer overnight at 4 °C. The following primary antibodies were used: anti-GFP (rabbit, 1:1,000, Molecular Probes), anti-GFP (chicken, 1:4,000, abcam), anti-K8 (rat, 1:500, Developmental Studies Hybridoma Bank), anti-K14 (rabbit, 1:2,000, Covance), anti-K14 (chicken, 1:1,000, Covance), anti-K5 (rabbit, 1:1,000, Covance), anti-K19 (rat, 1:500, Developmental Studies Hybridoma Bank), Cy3-coupled-anti-SMA (mouse, 1/1,000, Sigma), NaPiIIb (rabbit, 1/300, provided by J. Biber). The following secondary antibodies were used: anti-rabbit, anti-rat, anti-chicken conjugated to AlexaFluor 488 (Molecular Probes), to rhodamine Red-X or to Cy5 (JacksonImmunoResearch). Nuclei were stained in Hoechst solution and

slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Microscope image acquisition. For whole-mount, freshly dissected mammary glands were squeezed between two coverslips and were imaged using a Zeiss EC Plan-NEOFLUAR $\times 2.5$ objective. Pictures of immunostaining were acquired using $\times 10$ and $\times 40$ Zeiss EC Plan-NEOFLUAR objectives. Pictures were acquired on an Axio Observer Z1 microscope, with the AxioCamMR3 camera and using the Axiovision software (Carl Zeiss). Confocal pictures were acquired at room temperature using a Zeiss LSM780 multiphoton confocal microscope fitted on an Axiovert M200 inverted microscope equipped with C-Apochromat ($\times 40 = 1.2$ numerical aperture) water immersion objectives (Carl Zeiss). Optical sections, $1,024 \times 1,024$ pixels, were collected sequentially for each fluorochrome. The data sets generated were merged and displayed with the ZEN software.

Mammary cell preparation. Mammary glands were dissected and the lymph nodes removed before processing. Tissues were washed in HBSS, and cut in pieces of 1 mm³ with scissors. The tissues were placed in HBSS plus 300 U ml⁻¹ collagenase (Sigma) plus 300 μ g ml⁻¹ hyaluronidase (Sigma) and digested for 2 h at 37 °C under shaking. EDTA at a final concentration of 5 mM was added for 10 min to the resultant organoid suspension, followed by 0.25% Trypsin-EGTA for 2 min before filtration through a 40- μ m mesh, two successive washes in 2% FBS/PBS and labelling.

Cell labelling, flow cytometry and sorting. Two-five million cells per condition were incubated in 250 μ l of 2% FBS/PBS with primary antibodies for 30 min, with shaking every 10 min. Primary antibodies were washed with 2% FBS/PBS, and cells incubated for 30 min with the appropriate secondary antibodies, with shaking every 10 min. Secondary antibodies were washed with 2% FBS/PBS and cells were resuspended in 2.5 μ g ml⁻¹ DAPI (Invitrogen) before analysis. Primary antibodies used were: biotinylated anti-CD24 (1/50, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1/50, clone eBioHMB1-1, eBiosciences), PE-conjugated anti-CD45 (1/50, clone 30-F11, eBiosciences), PE-conjugated anti-CD31 (1/50, clone MEC 13.33, BD Biosciences), PE-conjugated anti-CD140a (1/50, clone APA5, eBiosciences). PE-Cy7 coupled Streptavidin (1/400, BD Biosciences) was used as secondary antibody. Data analysis and cell sorting was performed on a FACSaria sorter using the FACS DiVa software (BD Biosciences). Dead cells were excluded with DAPI; CD45⁺, CD31⁺ and CD140a⁺ cells were excluded (Lin⁻) before analysis of the YFP⁺ cells. For Lgr5-GFP-CreER/Rosa-Tomato and K14-Cre/Rosa-Tomato mice, the same procedure was followed, except that no Lin⁻ exclusion was performed. A minimum of 1,000 YFP⁺ cells was analysed per mouse.

Mammary fat-pad transplantation and analysis. One-million unsorted mammary cells (around 5×10^4 MECs) (for primary grafts), 10^5 unsorted cells from the primary graft (for secondary grafts) or 10,000 sorted cells (either only YFP⁺, or a combination of luminal Tomato⁺ and myoepithelial YFP⁺) as indicated were resuspended in 10 μ l PBS and injected into the number 4 glands of 4-week-old NOD/SCID female mice that had been cleared of endogenous epithelium. Recipient mice were mated 4 weeks after the transplantation, and were killed 3 weeks later, when usually fully pregnant. Recipient glands were dissected, analysed by whole-mount, fixed and placed in OCT for further staining. An outgrowth was defined as an epithelial structure comprising ducts and lobules and/or terminal end buds. The percentage of YFP⁺ luminal or myoepithelial cells was analysed on a minimum of 1,000 YFP⁺ cells on an average of 12 sections per graft. For the grafts of unsorted cells from K8-CreER/Rosa-YFP mice previously treated with 15 mg tamoxifen, a total of 7,885 and 6,591 YFP⁺ cells were analysed for the primary and secondary transplants, respectively. For the grafts of unsorted cells from K14-rtTA/TetO-Cre/Rosa-YFP mice previously treated with doxycycline food for 1 week, a total of 9,352 and of 5,505 YFP⁺ cells were analysed for the primary and secondary transplants, respectively. For the graft of mixed YFP⁺ and Tomato⁺ sorted cells, a total of 4,714 luminal cells was analysed for expression of YFP or Tomato.

Quantification of keratin expression in YFP⁺ cells. A minimum of 20 sections and about 10,000 YFP⁺ cells per mouse were analysed for K5 and K8 expression. For lineage tracing of K14-rtTA/TetO-Cre/Rosa-YFP, a total of 32,853, 31,429, 31,633 and 34,002 cells from three different mice per condition were analysed respectively at 1 week and 10 weeks after tamoxifen administration at 4 weeks, and during first lactation and second lactation. No K8-expressing cells were observed among the YFP⁺ cells in these conditions. For lineage tracing of K8-CreER/Rosa-YFP, a minimum of 10,000 YFP⁺ cells per mouse from three different mice per condition were analysed respectively at 1 week and 10 weeks after tamoxifen administration at puberty, during first lactation and during second lactation. No K5-expressing cells were observed among the YFP⁺ cells in these conditions.

Quantification of the number and size of clones induced in the mammary gland. A minimum of 500 YFP⁺ clones were analysed per time point for each CreER/Rosa-YFP lineage in three independent mice. For each mammary tube section, the number, size (number of cells) and keratin expression (myoepithelial (K5) or luminal (K8)) of YFP clones was scored. The clones were grouped in three

classes according to their size, and their frequency was calculated as the percentage of the total number of clones. The mean number of clones per 1,000 cells was calculated at the different time points studied and normalized as the percentage of the number of clones counted 1 week after tamoxifen administration. For the confocal analysis of size of clones in K8-CreER/Rosa-YFP mice, 69 and 197 clones were analysed respectively 1 week and at 4 weeks after tamoxifen induction.

33. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
34. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
35. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of α -catenin in skin. *Cell* **104**, 605–617 (2001).
36. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
37. Perl, A. K., Wert, S. E., Nagy, A., Lobe, C. G. & Whitsett, J. A. Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl Acad. Sci. USA* **99**, 10482–10487 (2002).

Structure and function of the AAA⁺ protein CbbX, a red-type Rubisco activase

Oliver Mueller-Cajar¹, Mathias Stotz¹, Petra Wendler², F. Ulrich Hartl¹, Andreas Bracher¹ & Manajit Hayer-Hartl¹

Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyses the fixation of atmospheric CO₂ in photosynthesis, but tends to form inactive complexes with its substrate ribulose 1,5-bisphosphate (RuBP). In plants, Rubisco is reactivated by the AAA⁺ (ATPases associated with various cellular activities) protein Rubisco activase (Rca), but no such protein is known for the Rubisco of red algae. Here we identify the protein CbbX as an activase of red-type Rubisco. The 3.0-Å crystal structure of unassembled CbbX from *Rhodobacter sphaeroides* revealed an AAA⁺ protein architecture. Electron microscopy and biochemical analysis showed that ATP and RuBP must bind to convert CbbX into functionally active, hexameric rings. The CbbX ATPase is strongly stimulated by RuBP and Rubisco. Mutational analysis suggests that CbbX functions by transiently pulling the carboxy-terminal peptide of the Rubisco large subunit into the hexamer pore, resulting in the release of the inhibitory RuBP. Understanding Rubisco activation may facilitate efforts to improve CO₂ uptake and biomass production by photosynthetic organisms.

The enzyme Rubisco is responsible for the entry of inorganic carbon into the biosphere, catalysing the fixation of atmospheric CO₂ by carboxylation of RuBP in photosynthesis^{1,2}. The major form of Rubisco (form I) is hexadecameric, consisting of eight large (RbcL) and eight small (RbcS) subunits. The form I Rubiscos are phylogenetically divided into a green branch, present in cyanobacteria, green algae and plants, and a red branch, present mainly in photosynthetic bacteria, red algae and phytoplankton^{3–5}. The red-type Rubiscos are responsible for most oceanic CO₂ uptake, and many display higher CO₂/O₂ specificity values than their green-type counterparts^{6,7}.

To become catalytically active, form I Rubisco must first be carbamylated by a non-substrate CO₂ molecule at the active-site lysine and bind Mg²⁺ as cofactor⁸ (Fig. 1a). Premature binding of the substrate, RuBP, to uncarbamylated Rubisco results in an inactive complex⁹. Green algae and all plants possess Rubisco activase (Rca), an AAA⁺ protein that catalyses the release of RuBP from inhibited Rubisco in an ATP-dependent manner¹⁰ (Fig. 1a). AAA⁺ proteins generally function in protein unfolding and disassembly¹¹, but the mechanism of Rubisco activation has remained enigmatic¹⁰.

In view of the importance of Rca for Rubisco function, it is surprising that the genomes of organisms containing red-type Rubisco apparently encode no Rca homologue¹². However, the genes encoding the Rubisco subunits are frequently associated with a gene called *cbbX*, encoding a protein of about 35 kDa (refs 13, 14). CbbX has been proposed to be involved in the transcriptional regulation of the Rubisco operon¹⁵. Here we report the crystal structure of CbbX from the proteobacterium *Rhodobacter sphaeroides* and show that it is an AAA⁺ protein that functions as an activase for red-type Rubisco.

CbbX is a Rubisco activase

Inactivation of *cbbX* results in impaired photoautotrophic growth in *R. sphaeroides* (Rs) and other proteobacteria^{13,16}. However, co-expression of RsCbbX and RsRubisco in *Escherichia coli* did not improve the yield of soluble Rubisco hexadecamer, suggesting that CbbX does not function as an assembly chaperone analogous to RbcX in cyanobacteria^{17,18} (data not shown). A BLAST homology search suggested that CbbX belongs to the AAA⁺ protein family; we therefore investigated the possibility that CbbX

is a red-type Rubisco activase. Purified recombinant RsRubisco, activated by carbamylation and Mg²⁺ binding (E.C.M), fixed CO₂ at a linear rate of about 2.0 s⁻¹ (Fig. 1b). In contrast, incubation of Rubisco with RuBP in the absence of Mg²⁺ generated the inhibited enzyme (E.I), which showed negligible CO₂ fixation (Fig. 1b), as reported for the native, non-recombinant RsRubisco¹⁹. The inhibited Rubisco was efficiently activated in the presence of CbbX and ATP (Fig. 1b). Activation required ATP hydrolysis, because ATP could not be replaced by ADP or non-hydrolysable ATP analogues (data not shown). The speed of activation was dependent on the concentration of CbbX (Fig. 1b); at 5 μM CbbX protomer, 2 μM Rubisco (protomer) reached full activity within 150 s.

CbbX alone showed no detectable ATPase activity, in contrast to Rca, which is constitutively ATPase active²⁰ (Supplementary Table 1a). The addition of RuBP resulted in a concentration-dependent activation of the CbbX ATPase, with half-maximal rate at about 35 μM RuBP (Fig. 1c and Supplementary Table 1a). The structurally similar pentose sugars D-ribulose 5-phosphate and D-ribose 5-phosphate failed to stimulate the ATPase (data not shown). However, the maximum ATPase rate (8.0 ± 0.2 min⁻¹) in the presence of RuBP was only about 20% of that reported for Rca²⁰. The addition of inhibited Rubisco (E.I) stimulated the ATPase rate of CbbX up to about tenfold with an apparent K_d of about 3 μM (Fig. 1d). A somewhat lower ATPase stimulation was also measured with active Rubisco (E.C.M) (Fig. 1d), suggesting that CbbX distinguishes only poorly between E.I and E.C.M, at least *in vitro*. Indeed, published crystal structures indicate that the active and inhibited Rubisco complexes have virtually indistinguishable surface properties^{21,22}. The stimulating effect by E.I and E.C.M was observed only in the presence of free RuBP, reflecting conditions of active photosynthesis²³ (Supplementary Table 1a). Thus, full stimulation of the CbbX ATPase for Rubisco activation requires the binding of RuBP to CbbX and the simultaneous recognition of Rubisco. This regulation would ensure that CbbX is active when photosynthesis is ongoing and Rubisco is at risk of inactivation (Fig. 1a).

Oligomeric state of functionally active CbbX

Purified CbbX forms oligomeric assemblies of various sizes. To determine the functionally active oligomeric state of CbbX, we employed

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany. ²Gene Center Munich, Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.

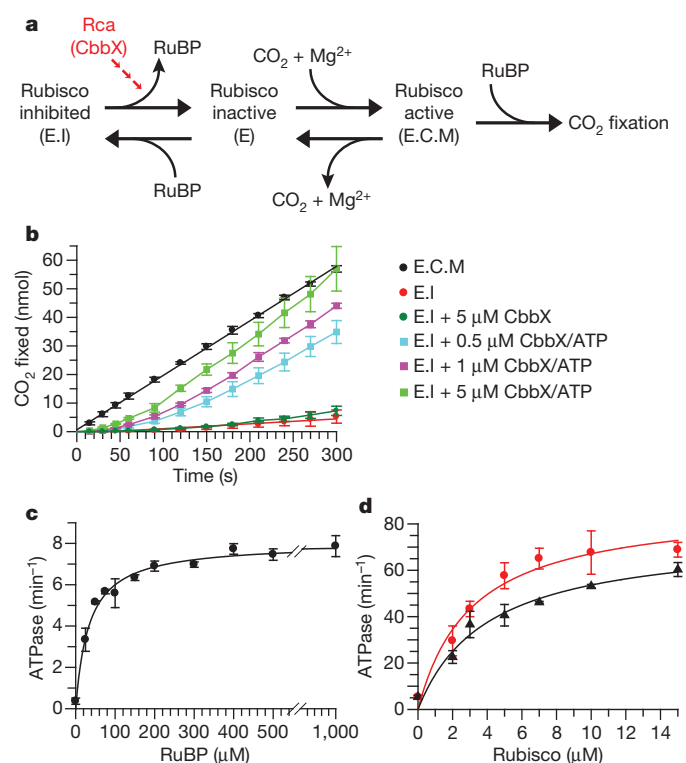


Figure 1 | CbbX functions as a Rubisco activase. **a**, Schematic representation of form I Rubisco activation and inhibition (modified from ref. 38). Uncarbamylated, inactive enzyme E can bind RuBP to form the inhibited complex E.I or it can react with CO₂ and Mg²⁺ to form the active E.C.M complex. Rca (CbbX) disrupts E.I and releases the inhibition. **b**, CO₂ fixation assays (50-μl reactions) were performed with E.C.M or E.I Rubisco (2 μM protomer) in the absence or presence of CbbX (0.5–5 μM protomer) with 4 mM ATP when indicated. **c**, Dependence of CbbX ATPase on RuBP concentration. ATPase activity of CbbX (5 μM protomer) was measured at increasing concentrations of RuBP. $K_d = 34 \pm 3.3 \mu\text{M}$; $V_{\text{max}} = 8.0 \pm 0.2 \text{ min}^{-1}$. **d**, Dependence of CbbX ATPase on Rubisco concentration. ATPase activity of CbbX (1 μM protomer) was measured in the presence of 1 mM RuBP and increasing concentrations of E.I Rubisco (red; $K_d = 3.1 \pm 0.8 \mu\text{M}$; $V_{\text{max}} = 88 \pm 7.1 \text{ min}^{-1}$) or E.C.M Rubisco (black; $K_d = 4.0 \pm 1.0 \mu\text{M}$; $V_{\text{max}} = 75 \pm 6.6 \text{ min}^{-1}$). Error bars represent s.d. for at least three independent experiments.

negative-stain electron microscopy (EM) and multi-angle light scattering. CbbX alone formed amorphous particles of about 600–900 kDa (Fig. 2a and data not shown). In the presence of ATP (or non-hydrolysable nucleotide), long fibrillar structures were observed (Fig. 2b), consistent with a molecular mass of about 5–10 MDa (data not shown). In the presence of both ATP and RuBP, but not RuBP alone, CbbX formed ring structures (Fig. 2c, d). Class averages of these particles showed that CbbX forms a six-fold symmetric ring with a diameter of about 140 Å, a height of about 45 Å and a central pore with a diameter of about 25 Å, similar to the dimensions of other AAA⁺ proteins^{24–26} (Fig. 2e and Supplementary Fig. 1a–c). Eigenimage analysis of top views confirmed the six-fold symmetry (Supplementary Fig. 1d).

Crystal structure of CbbX

The crystal structure of *RsCbbX* was solved by selenium-single-wavelength anomalous dispersion (Se-SAD) at 3.1 Å resolution. The model was built against isomorphous native data at 3.0 Å resolution and refined to final *R* and *R*_{free} values of 0.218 and 0.285, respectively (Supplementary Table 2). The asymmetric crystal unit contains two copies of CbbX (Supplementary Fig. 2a). CbbX has a typical AAA⁺ family protein architecture¹¹, containing an amino-terminal α/β subdomain (residues 34–205) and a smaller carboxy-terminal α-helical

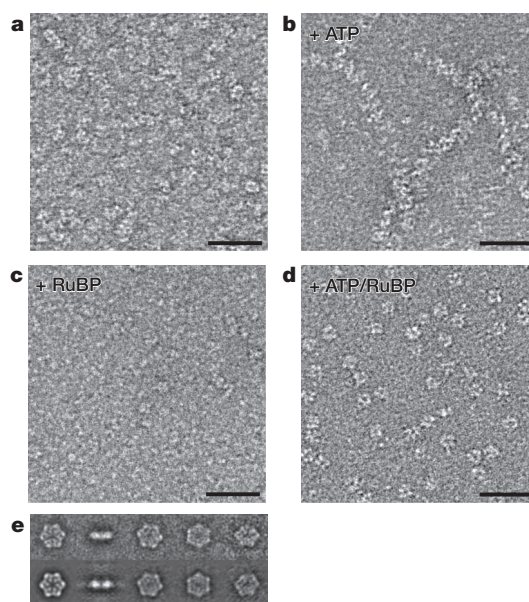


Figure 2 | Negative-stain electron microscopy of CbbX. **a–d**, CbbX (0.1 mg ml⁻¹) was incubated for 5 min at 25 °C either in buffer (**a**) or in buffer containing 1 mM MgATP (**b**), 1 mM RuBP (**c**) or MgATP and RuBP (**d**). Scale bars, 50 nm. **e**, The upper panel shows class-average images of CbbX incubated with MgATP, ATP-γS and RuBP (1 mM each) obtained by multivariate statistical analysis. The lower panel shows corresponding re-projections of the final three-dimensional map in the Euler angle direction assigned to the class averages. Each class average contains 15–20 images.

subdomain (residues 206–296), separated by a short linker (Fig. 3a). The N-terminal domain contains the canonical Walker A and B motifs (residues 74–81 and 121–138, respectively) and a pore loop with the sequence Tyr-Ile-Gly (residues 114–116; Fig. 3a and Supplementary Fig. 3). The Walker A and B mutants K80A and E138Q, respectively, were inactive with regard to ATP hydrolysis and Rubisco activation (Supplementary Table 1b). The α/β subdomain has a conserved N-terminal extension (residues 8–33) that is unique to CbbX (Fig. 3a and Supplementary Fig. 3). It contains two α-helices, α1 and α2, of which the latter forms a three-helix bundle with helices α3 and α4 of the α/β subdomain (Fig. 3a). The N-terminal residues 1–7 are probably flexible, because they are structured in only one CbbX molecule in the asymmetric unit. Each CbbX chain contains three bound sulphate ions from the precipitant, one in the ATP-binding site and two in the α-helical subdomain (Fig. 3a).

The crystal contains two slightly different conformations of CbbX differing by 13° in the orientation of the α-helical subdomain (Supplementary Fig. 2b). This suggests that the inter-domain linker functions as a flexible hinge, while the subdomains behave as rigid units with root mean squared deviation values of 0.703 Å (α/β subdomain) and 0.390 Å (α-helical subdomain), respectively. The α-helical subdomain (helices α10 and α11) of each CbbX chain is involved in an extensive contact with helices α1 and α3 of the α/β subdomain of its neighbouring CbbX chain, burying roughly 770 Å² on each partner (Fig. 3b). Two virtually identical copies of this contact were found in the crystal lattice (Supplementary Fig. 2c–e). Other crystal contacts are unique, providing no indication for ordered oligomer structure. Tight helix–helix packing at the inter-subunit interface is enabled by the small, opposed alanine residues 48 and 263 in helices α3 and α10, respectively (Fig. 3b), which are highly conserved (Supplementary Fig. 3). These alanines are surrounded by conserved hydrophobic residues (Fig. 3b and Supplementary Fig. 3). The mutant A48N has poor activase function (about 12% residual activity; Supplementary Table 1d); it forms relatively few, short fibrillar structures in the presence of ATP and no stable hexamers in the presence of ATP and RuBP (data not shown). Thus, the extensive subunit interface

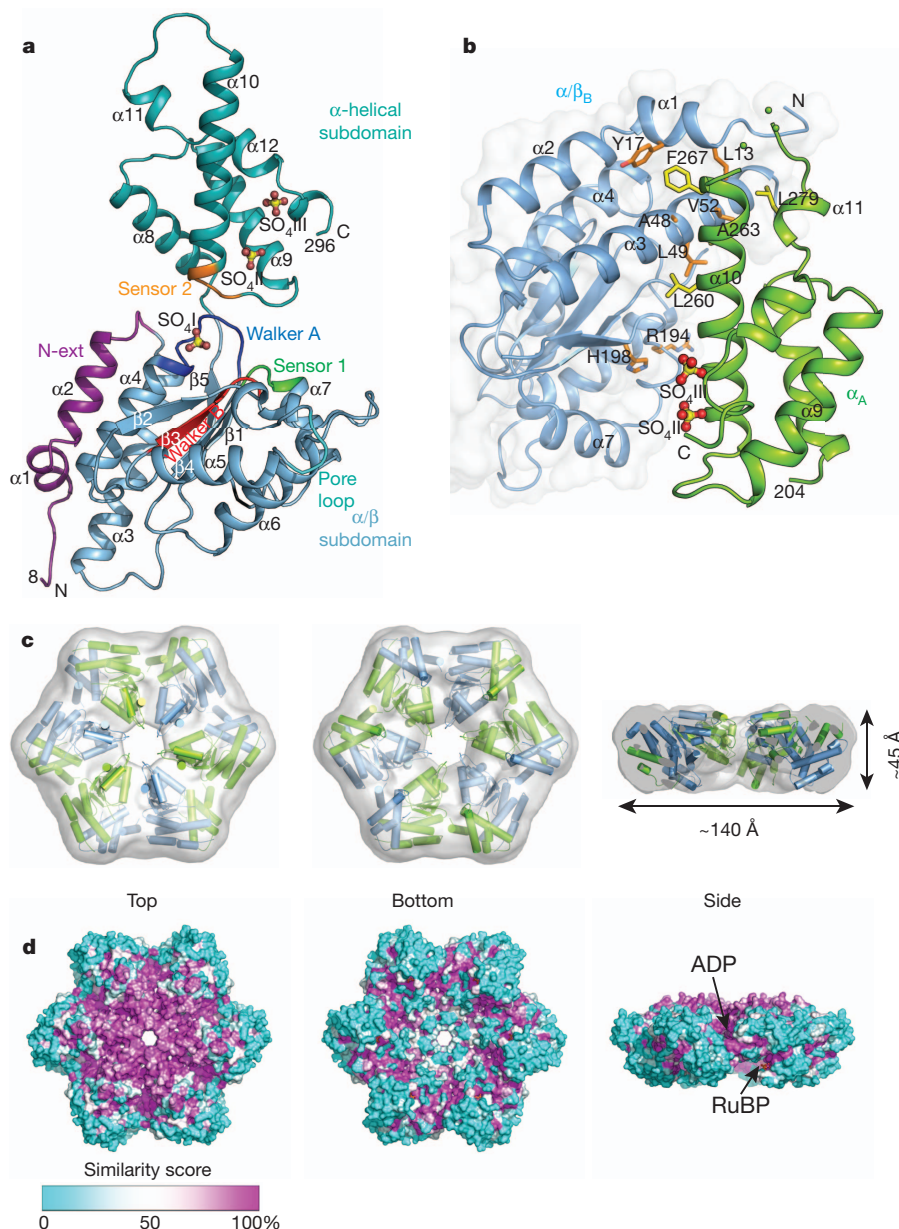


Figure 3 | Crystal structure and hexamer model of CbbX. **a**, Ribbon representation of the crystal structure of *R. sphaeroides* CbbX. The α/β and the α -helical subdomains are indicated in blue and teal, respectively, and the N-terminal extension in purple. The model shown is a composite of the two CbbX chains in the asymmetric unit. The Walker A and B motifs are shown in dark blue and red, and the sensor I and II regions in green and orange, respectively. The pore loop is indicated. Bound sulphates are shown in ball-and-stick form. N-ext, N-terminal extension. **b**, The CbbX-CbbX contact interface in the asymmetric unit. The α -helical subdomain of one CbbX chain

observed in the crystal structure is likely to prevail in the functional CbbX hexamer.

Hexameric ring structure of CbbX

To understand the structural features of CbbX in the context of an active hexameric particle (Fig. 2d, e), we used the module consisting of the α/β subdomain of one CbbX chain bound to the α -helical subdomain of the adjacent CbbX chain (Fig. 3b) and modelled it onto the ATPase-active D2 ring of the hexameric p97 AAA⁺ complex (Protein Data Bank accession code 3CF3)²⁷. This module was proposed to form a building block that is invariant through the conformational changes of AAA⁺ hexamers²⁸. The resulting CbbX hexamer is well accommodated by the six-fold symmetric density under the electron microscope (Fig. 3c). The

(α_A , green) contacts helices $\alpha 1$ and $\alpha 3$ of the α/β subdomain of the adjacent CbbX chain (α/β_B , blue). Conserved interface residues are indicated. **c**, Overlay of a hexamer model based on six α_A - α/β_B units (as shown in **b**) onto the three-dimensional electron microscope reconstruction of the CbbX hexamer (Fig. 2d, e). The subunits are indicated alternately in green and blue. **d**, Surface conservation of the CbbX hexamer, based on an alignment of 62 CbbX sequences, mapped onto the surface of the CbbX hexamer model. The positions of ADP and RuBP are indicated.

last residues of the α/β subdomain are positioned within binding distance of the first residues of the α -helical subdomain of the same chain without imposing any constraints. In comparison with its conformation in the crystal, the α -helical subdomain is shifted and rotated about 46° towards the α/β subdomain (Fig. 3c and Supplementary Fig. 4a, b). This reorientation is necessary to accommodate ATP in the putative nucleotide-binding pocket (Fig. 3a and Supplementary Fig. 4c). Moreover, it positions the residue Arg 194 as the putative arginine finger¹¹ close to the nucleotide-binding site of the subsequent subunit (Fig. 3b). Mutation of Arg 194 to alanine resulted in a loss of ATPase and activase function (Supplementary Table 1b). CbbX(R194A) formed fibrillar structures in the presence of ATP but failed to form hexamers in the presence of ATP and RuBP (Supplementary Fig. 4d).

To locate functionally important surface regions, the sequence conservation score from an extensive alignment of CbbX sequences was plotted onto the CbbX hexamer model. As expected, surface conservation is high at the subunit interfaces (Fig. 3d and Supplementary Fig. 4c). The top surface of the hexamer close to the central pore, including the pore loop and helices η 1, α 5 and α 6, is also highly conserved (Fig. 3d and Supplementary Fig. 3). In contrast, the bottom surface of the hexamer shows poor surface conservation, except for a region in the α -helical subdomain (Fig. 3d and Supplementary Fig. 4c) that represents the putative RuBP-binding pocket (see Fig. 4a, b). The conserved surfaces of the α/β subdomains at the hexamer top (Fig. 3d) probably harbour the interaction sites with Rubisco.

Allosteric regulation by RuBP

A feature of the α -helical subdomain of CbbX in the crystal is the presence of two sulphate ions, separated by only about 8 Å, which are bound in a positively charged pocket (Fig. 4a, b). A third sulphate is bound at the Walker A motif in the α/β subdomain (Fig. 3a). The distance of the sulphate ions in the α -helical subdomain approximately fits the phosphate-phosphate distance in RuBP, which is 9.3 Å in the fully extended conformation; RuBP can be modelled into the pocket, suggesting that the bound sulphates mark the position of the allosteric RuBP-binding site (Fig. 4b). Consistent with this possibility, the stimulating effect of RuBP on the CbbX ATPase was inhibited by about 40% by 50 mM ammonium sulphate and completely inhibited by 200 mM ammonium sulphate (Supplementary Table 1a). Furthermore, negative-stain electron microscopy showed that in the presence of ATP, RuBP and excess sulphate, CbbX formed fibrillar structures, suggesting that the sulphate ions prevented the RuBP-induced formation of hexamers (Fig. 2d and Supplementary Fig. 5a). The two sulphate ions are contacted by the highly conserved residues Arg 239, Ser 250, Asn 253, Arg 257 and Arg 261 (Fig. 4a and Supplementary Fig. 3), which are located adjacent to the inter-subunit interface of the hexamer (Fig. 3b). Individual mutations of these residues were deficient for ATP hydrolysis in the presence of RuBP and were unable to activate the inhibited Rubisco complex (Supplementary Table 1c). Almost all mutants formed fibrils in the presence of ATP and RuBP; the exception was CbbX(N253D), which formed amorphous particles (Supplementary Fig. 5b).

The highly conserved residue His 198 faces the putative RuBP-binding site in the adjacent CbbX subunit (Fig. 3b) and may thus contribute to RuBP binding and CbbX allostery. Indeed, the mutation H198F, which would be compatible with the hydrophobic vicinity of His 198, strongly inhibited the CbbX ATPase and activase functions (Supplementary Table 1c). The side chain of His 198 would have to rotate almost 180° in the hexamer model compared with the crystal structure to contact the carbohydrate moiety of RuBP, but this movement would be possible without steric hindrance (Fig. 3b). The putative arginine finger residue Arg 194 is located in close proximity to His 198 and might be positioned in the ATP-hydrolysis active state on binding of RuBP. Consistent with the existence of a regulatory network involving Arg 194, His 198 and RuBP, the mutation H198F resulted in a loss of hexamer formation in the presence of ATP and RuBP, as was observed with R194A (Supplementary Fig. 4d and data not shown).

Mechanism of Rubisco activation

Studies on other protein-remodelling AAA⁺ ATPases have revealed the importance of a loop region facing the central pore of the hexamer, containing a highly conserved Tyr(Val/Ile)Gly motif, which is involved in threading substrate polypeptide into or through the pore^{11,29–32}. This motif is also conserved in CbbX (Supplementary Fig. 3) and is facing the pore in the hexamer model (Fig. 4c).

To test whether this loop is also important for CbbX function, we purified and characterized the Y114A mutant. The ATPase rate of the Y114A mutant, when stimulated with RuBP, was almost double that of

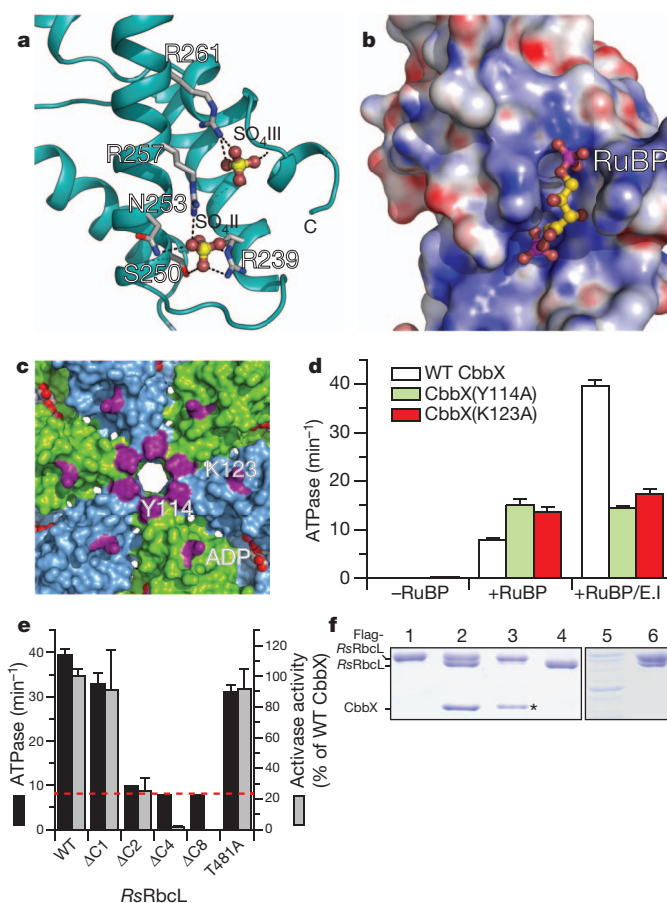


Figure 4 | Structural and functional analysis of CbbX mechanism.

a, Putative RuBP-binding pocket on CbbX. The two sulphate ions in the α -helical subdomain are shown in ball-and-stick form, and key contact residues in stick representation. Nitrogen, oxygen and sulphur atoms are shown in blue, red and yellow, respectively. Hydrogen-bond distances are indicated by dotted lines. **b**, Surface representation of the RuBP pocket. Colour gradient from red to blue indicates -30 to $+30 k_B T$ electrostatic surface potential. RuBP was modelled into the pocket. **c**, Closer view of the top surface of the hexamer model with bound ADP, showing the location of the conserved Lys 123 and pore residue Tyr 114 (purple). **d**, ATPase activity of the CbbX mutants Y114A and K123A in the absence or presence of RuBP or RuBP/E.I. (Supplementary Table 1d). WT, wild-type. **e**, CbbX ATPase and activase activity with C-terminally truncated RsrBcL (Δ C1 to Δ C8) and with RbcL(T481A). Dotted red line, ATPase activity with RuBP alone. Assay conditions as in Supplementary Table 1. **f**, Rubisco E.I. does not disassemble during activation. Left: a 1:1 mixture of Flag-tagged and untagged Rubisco E.I. was incubated for activation (10 min, 25 °C) (lane 2) and subjected to anti-Flag pulldown (lane 3). Flag-tagged and untagged Rubisco are shown in lanes 1 and 4, respectively. The asterisk indicates that CbbX binds to the beads non-specifically. Right: the soluble lysate of *E. coli* cells co-expressing Flag-Rubisco and untagged Rubisco (lane 5) was subjected to anti-Flag pulldown (lane 6). See Methods for experimental details.

wild-type CbbX (Fig. 4d), but no activase activity was measurable (Supplementary Table 1d). This functional defect was correlated with the inability of E.I. Rubisco to stimulate the ATPase rate of CbbX(Y114A) further, in contrast to the effect observed with wild-type CbbX (Fig. 4d). In view of the strong conservation of the top surface of the CbbX hexamer (Fig. 3d), we speculated that the highly conserved residue Lys 123 (Supplementary Fig. 3) may also be involved in interacting with Rubisco (Fig. 4c). Similarly to the Y114A mutation, the K123A mutant had an increased ATPase rate in the presence of RuBP but was not stimulated further by E.I. or E.C.M. Rubisco (Supplementary Table 1d). The K123A mutant had a low residual activase function of about 18% (Supplementary Table 1d). These results demonstrate the functional importance of the pore region of CbbX

and suggest that mutation of this region uncouples the CbbX ATPase and activase functions.

In considering the sequence element(s) of Rubisco that might be recognized by the pore region of CbbX, we noted that red-type form I RbcL subunits possess a conserved, flexible C-terminal extension of about ten residues not present in the green-type subunits (Supplementary Fig. 5c). In the inactive, closed conformation of form I Rubisco, the preceding segment is packed against the catalytically important loop 6, which traps the bound RuBP^{2,33}. Exerting a pulling force on the accessible C terminus would destabilize the closed conformation, resulting in opening of the active site and release of the inhibitor. To test this hypothesis, we constructed C-terminally truncated Rubisco mutants (*RsrbcL* Δ C1 to Δ C8) and the alanine mutation of the highly conserved Thr 481. All mutants formed RbcL₈S₈ holoenzyme complexes with wild-type activity (data not shown). In contrast to the wild-type enzyme, Rubisco Δ C8, Δ C4 and Δ C2 did not stimulate the CbbX ATPase activity above the value measured in the presence of RuBP alone, whereas Δ C1 and T481A preserved this function (Fig. 4e). Thus, engagement of the Rubisco C-terminal tail by the central pore of CbbX is necessary for ATPase stimulation. Accordingly, the inhibited complexes of Δ C8 and Δ C4 Rubisco could not be activated by CbbX, whereas Δ C2 was activated with about 20% efficiency and Δ C1 and T481A were fully activated (Fig. 4e).

To investigate whether activation involves disassembly of the Rubisco holoenzyme, we performed experiments with a mixture of inhibited Rubisco complexes consisting of either N-terminally Flag-tagged or untagged RbcL subunits. During activation by CbbX, no Rubisco complexes containing both types of subunit were formed (Fig. 4f, lane 3), although such complexes were observed on co-expression of tagged and untagged subunits in *E. coli* (Fig. 4f, lane 6). This finding indicates that activation does not involve Rubisco disassembly. CbbX instead remodels the Rubisco complex by transiently pulling on the RbcL C-terminal tail.

Conclusions

We have shown in this study that the AAA⁺ protein CbbX functions as an activase for red-type form I Rubisco under allosteric regulation

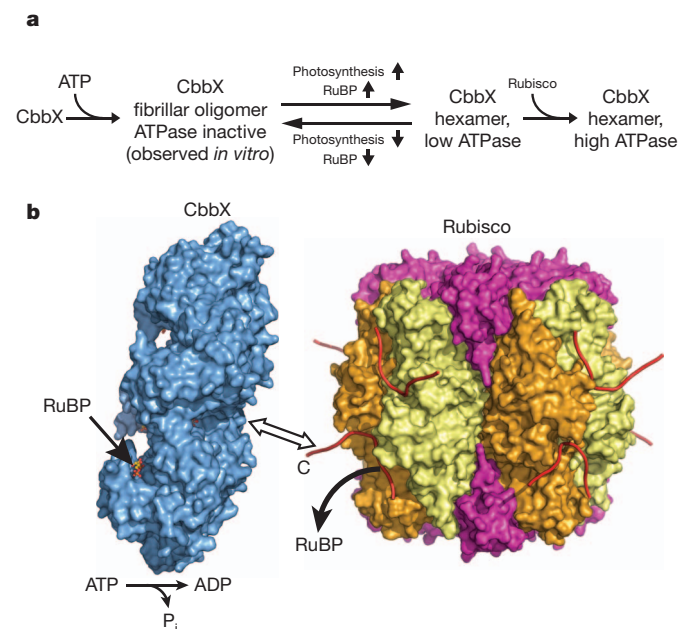


Figure 5 | Model of Rubisco activation by CbbX. **a**, Conformational and functional regulation of CbbX by RuBP and Rubisco (see the text). **b**, Proposed mechanism of Rubisco remodelling by CbbX. The surface-accessible C-terminal peptide of the RbcL subunit is transiently pulled into the central pore of CbbX, mediated by the CbbX ATPase, resulting in the release of inhibitory RuBP.

by RuBP, the substrate of the target protein. Furthermore, the CbbX ATPase rate is strongly stimulated by Rubisco. These features are not shared by the canonical Rubisco activase of higher plants^{10,20}. The biochemical and structural properties of CbbX suggest an intricate regulatory cycle to ensure efficient Rubisco function (Fig. 5a): in the absence of photosynthetic activity, the concentration of RuBP is expected to be very low and CbbX is inactive, perhaps populating the fibrillar assembly observed *in vitro* and thereby avoiding unnecessary ATP consumption. Activation of photosynthesis results in the accumulation of free RuBP, possibly reaching millimolar concentration²³. The free RuBP binds to CbbX, inducing its rearrangement to catalytically competent hexamers, which recognize the inhibited (and active) Rubisco complexes.

Docking of CbbX onto Rubisco probably involves the highly conserved top surface of the hexamer. Our mutational analysis further suggests that the accessible C-terminal tail of the Rubisco large subunit is engaged by the conserved pore loop of CbbX, resulting in stimulation of the CbbX ATPase. The C-terminal tail is probably transiently pulled into the hexamer pore (Fig. 5b), triggering disruption of the closed Rubisco-inhibitor complex. Although the details of this mechanism remain to be investigated, our results indicate that activation does not involve complete threading of RbcL subunits through the central pore of CbbX. It will also be interesting to see whether the activation mechanism is cooperative within the antiparallel RbcL dimers or the holoenzyme, or whether each active site must be remodelled individually.

METHODS SUMMARY

Proteins. The *cbbL/cbbS* and *cbbX* genes were amplified from genomic DNA of *R. sphaeroides* 2.4.1 and cloned into pET30b (Novagen) and into the pHue vector³⁴, respectively. Rubisco and CbbX proteins were expressed in *E. coli* and purified as described in Methods and in ref. 35.

Enzymatic assays. CO₂ fixation was measured at 25 °C in 50 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 30 mM NaH¹⁴CO₃ (25 Bq nmol⁻¹), 3 mM RuBP, 4 mM ATP and Rubisco E.C.M, E.I and CbbX as stated in figure legends. Rubisco E.C.M, E.I and E.I in the absence of free RuBP were prepared as described in Methods. Relative CbbX activities were measured by determining the increase in Rubisco activity during the first minute of a CO₂ fixation assay³⁶. ATPase activity was assayed spectrophotometrically using a coupled assay that followed the oxidation of NADH³⁷.

Pulldowns from activation reactions containing a 1:1 mixture of Flag-tagged and untagged *Rsrubisco* were performed with EZview Red anti-Flag M2 Affinity gel (Sigma).

Electron microscopy and reconstruction. CbbX was negatively stained with 2% (w/v) uranyl acetate and analysed.

Crystallization and data collection. CbbX crystals were grown with 50 mM MES-NaOH pH 6.5 and 0.4 M ammonium sulphate as a precipitant. Diffraction data were collected at the European Synchrotron Radiation Facility, Grenoble, and the structure was solved by Se-SAD.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 April; accepted 15 September 2011.

Published online 2 November 2011.

1. Spreitzer, R. J. & Salvucci, M. E. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* **53**, 449–475 (2002).
2. Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008).
3. Tabita, F. R. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth. Res.* **60**, 1–28 (1999).
4. Tabita, F. R., Satagopan, S., Hanson, T. E., Kree, N. E. & Scott, S. S. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524 (2008).
5. Badger, M. R. & Bek, E. J. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J. Exp. Bot.* **59**, 1525–1541 (2008).
6. Whitney, S. M., Baldet, P., Hudson, G. S. & Andrews, T. J. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* **26**, 535–547 (2001).
7. Falkowski, P. G. *et al.* The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).

8. Lorimer, G. H., Badger, M. R. & Andrews, T. J. The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry* **15**, 529–536 (1976).
9. Jordan, D. B. & Chollet, R. Inhibition of ribulose bisphosphate carboxylase by substrate ribulose 1,5-bisphosphate. *J. Biol. Chem.* **258**, 13752–13758 (1983).
10. Portis, A. R. Jr. Rubisco activase—Rubisco's catalytic chaperone. *Photosynth. Res.* **75**, 11–27 (2003).
11. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
12. Pearce, F. G. Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem. J.* **399**, 525–534 (2006).
13. Gibson, J. L. & Tabita, F. R. Analysis of the *cbbXYZ* operon in *Rhodospirillum rubrum*. *J. Bacteriol.* **179**, 663–669 (1997).
14. Maier, U. G., Fraunholz, M., Zauner, S., Penny, S. & Douglas, S. A nucleomorph-encoded CbbX and the phylogeny of RuBisCo regulators. *Mol. Biol. Evol.* **17**, 576–583 (2000).
15. Fujita, K., Tanaka, K., Sadaie, Y. & Ohta, N. Functional analysis of the plastid and nuclear encoded CbbX proteins of *Cyanidioschyzon merolae*. *Genes Genet. Syst.* **83**, 135–142 (2008).
16. Bowien, B. & Kusian, B. Genetics and control of CO₂ assimilation in the chemolithotroph *Ralstonia eutropha*. *Arch. Microbiol.* **178**, 85–93 (2002).
17. Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
18. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
19. Gibson, J. L. & Tabita, F. R. Activation of ribulose 1,5-bisphosphate carboxylase from *Rhodospseudomonas sphaeroides*: probable role of the small subunit. *J. Bacteriol.* **140**, 1023–1027 (1979).
20. Robinson, S. P. & Portis, A. R. Jr. Adenosine triphosphate hydrolysis by purified rubisco activase. *Arch. Biochem. Biophys.* **268**, 93–99 (1989).
21. Sugawara, H. *et al.* Crystal structure of carboxylase reaction-oriented ribulose 1,5-bisphosphate carboxylase/oxygenase from a thermophilic red alga, *Galdieria partita*. *J. Biol. Chem.* **274**, 15655–15661 (1999).
22. Okano, Y. *et al.* X-ray structure of *Galdieria* Rubisco complexed with one sulfate ion per active site. *FEBS Lett.* **527**, 33–36 (2002).
23. Von Caemmerer, S. & Edmondson, D. L. Relationship between steady-state gas exchange *in vivo* ribulose bisphosphate carboxylase activity and some carbon reduction cycle intermediates in *Raphanus sativus*. *Aust. J. Plant Physiol.* **13**, 669–688 (1986).
24. Sousa, M. C. *et al.* Crystal and solution structures of an HslUV protease–chaperone complex. *Cell* **103**, 633–643 (2000).
25. Massey, T. H., Mercogliano, C. P., Yates, J., Sherratt, D. J. & Lowe, J. Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol. Cell* **23**, 457–469 (2006).
26. Matias, P. M., Gorynia, S., Donner, P. & Carrondo, M. A. Crystal structure of the human AAA⁺ protein RuvBL1. *J. Biol. Chem.* **281**, 38918–38929 (2006).
27. Davies, J. M., Brunger, A. T. & Weis, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16**, 715–726 (2008).
28. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A. & Sauer, R. T. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**, 744–756 (2009).
29. Weibezahn, J. *et al.* Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **119**, 653–665 (2004).
30. Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W. & Horwich, A. L. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121**, 1029–1041 (2005).
31. Martin, A., Baker, T. A. & Sauer, R. T. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nature Struct. Mol. Biol.* **15**, 1147–1151 (2008).
32. Roll-Mecak, A. & Vale, R. D. Structural basis of microtubule severing by the hereditary spastic paraplegia protein spastin. *Nature* **451**, 363–367 (2008).
33. Andersson, I. Catalysis and regulation in Rubisco. *J. Exp. Bot.* **59**, 1555–1568 (2008).
34. Catanzariti, A.-M., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* **13**, 1331–1339 (2004).
35. Baker, R. T. *et al.* Using deubiquitylating enzymes as research tools. *Methods Enzymol.* **398**, 540–554 (2005).
36. Esau, B. D., Snyder, G. W. & Portis, A. R. Jr. Differential effects of N- and C-terminal deletions on the two activities of rubisco activase. *Arch. Biochem. Biophys.* **326**, 100–105 (1996).
37. Kreuzer, K. N. & Jongeneel, C. V. *Escherichia coli* phage T4 topoisomerase. *Methods Enzymol.* **100**, 144–160 (1983).
38. Parry, M. A. J., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E. & Andralojc, P. J. Rubisco regulation: a role for inhibitors. *J. Exp. Bot.* **59**, 1569–1580 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Kaplan for providing the *R. sphaeroides* strain 2.4.1, S. Whitney for providing the pHue protein expression system, and R. Lange and N. Wischniewski for technical assistance. Support by the Max Planck Institute of Biochemistry (MPIB) Core Facility, the MPIB Crystallization Facility and the Joint Structural Biology Group staff at the European Synchrotron Radiation Facility beamlines is gratefully acknowledged. We thank the Deutsche Forschungsgemeinschaft (DFG) (SFB 594; DFG grant WE4628/1 to P.W.) and the Körber Foundation for financial support.

Author Contributions O.M.-C. designed and performed all the biochemical experiments. O.M.-C., M.S. and A.B. obtained the CbbX crystals and solved the structure. P.W. performed the electron microscopy and three-dimensional image analysis. All authors contributed to data interpretation and manuscript preparation. O.M.-C., A.B., F.U.H. and M.H.-H. wrote the manuscript.

Author Information Coordinates and structure factor amplitudes for CbbX crystal structures are deposited in the Protein Data Bank (PDB) under accession codes 3SYL and 3SYK; the hexamer model and the electron microscopy density are deposited in the PDB under accession code 3ZUH and in the Electron Microscopy Database (<http://www.ebi.ac.uk/pdbe/emdb/>) under accession code EMD-1932, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.H.-H. (mhartl@biochem.mpg.de) or A.B. (bracher@biochem.mpg.de).

METHODS

Plasmids. Genomic DNA of *Rhodobacter sphaeroides* 2.4.1 was prepared from cells grown in Luria–Bertani medium³⁹. The form I Rubisco genes *cbbL* and *cbbS* are positioned in tandem⁴⁰ and were amplified together and cloned between the *NdeI/HindIII* restriction sites of pET30b (Novagen) to yield pET30b*RscbbLS*. The *cbbX* gene was amplified and cloned between the *SacII/HindIII* restriction sites of pHue³⁴ to give pHue*RscbbX*. The 5' primer used was 5'-CTCCGCGGTGGTATGACCGACGCGGCAACGGC-3' (*SacII* site underlined, start codon italicized), allowing precise cleavage of the fused ubiquitin moiety during purification. The Quikchange protocol (Stratagene) was used to introduce point mutations and deletions into pET30b*RscbbLS* or pHue*RscbbX*. The Flag-tag sequence encoding MDYKDDDDKAA was inserted 5' to *RscbbL* in pET30b*RscbbLS* to give pET30bFLAG*RscbbLS*. The *XbaI/HindIII* fragments from pET30b*RscbbLS* and pET30bFLAG*RscbbLS* were cloned into pBAD18 and pBAD33 (ref. 41) to give pBAD18*RscbbLS* and pBAD33FLAG*RscbbLS*, respectively.

Protein expression and purification. All purification steps were performed at 4 °C and protein concentration was determined spectrophotometrically at 280 nm.

R. sphaeroides form I Rubisco was expressed in *E. coli* BL21 (DE3) cells, harbouring the plasmid pET30b*RscbbLS*, grown to an attenuation at 600 nm (D_{600}) of 0.5 at 37 °C in Luria–Bertani medium followed by induction for 4 h with 0.5 mM isopropyl β -D-thiogalactoside at 30 °C. For lysis, cells were incubated in 50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM EDTA, 0.5 mg ml⁻¹ lysozyme and Complete protease inhibitor cocktail (Roche) for 30 min on ice, followed by ultrasonication (Misonix Sonicator 3000). The supernatant obtained by high-speed centrifugation (48,000g, 45 min, 4 °C) was applied to a Source30Q column (Amersham Biosciences) equilibrated with 50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM EDTA, and the proteins were eluted with a linear NaCl gradient from 20 mM to 1 M. Rubisco-containing fractions were dialysed against 20 mM Tris-HCl pH 7.5, 20 mM NaCl, and applied to an equilibrated MonoQ column; proteins were eluted with a linear salt gradient to 0.5 M NaCl. The main fractions containing Rubisco activity were concentrated and applied to a Superdex 200 gel-filtration column equilibrated in buffer A (20 mM Tris-HCl pH 7.5, 50 mM NaCl). The purest fractions (more than 95% pure by SDS–PAGE) were concentrated, supplemented with 5% glycerol, flash-frozen in liquid N₂ and stored at –80 °C.

To produce Rubisco containing Flag-tagged RbcL or co-assembled Flag-tagged and untagged RbcL subunits, *E. coli* Top10 cells harbouring plasmids encoding untagged (pBAD18*RscbbLS*) and/or N-terminally Flag-tagged (pBAD33FLAG*RscbbLS*) RbcL subunits and untagged small subunits were induced during exponential phase at 30 °C for 3.5 h with 0.4% (w/v) L-arabinose. The cells were pelleted and after resuspension were lysed by ultrasonication in buffer B (50 mM Tris-HCl pH 8.0, 10 mM MgCl₂) containing 1 mM phenylmethylsulphonyl fluoride. Cellular debris was removed by centrifugation (16,100g, 4 °C, 15 min). Soluble lysate fractions were used in anti-Flag pull-downs (see below).

R. sphaeroides CbbX was produced as a His₆-ubiquitin fusion by using the pHue vector system and captured by immobilized metal-ion affinity chromatography (IMAC) followed by cleavage of the His₆-ubiquitin moiety to give the native N terminus, as described for other proteins³⁵. The His₆-ubiquitin moiety was cleaved overnight at 23 °C using the deubiquitinating enzyme Usp2 (ref. 35). The protein solution was dialysed against buffer A and applied to a MonoQ column equilibrated with buffer A; proteins were eluted with a linear salt gradient to 1 M NaCl. Fractions containing CbbX were combined and concentrated; 5% glycerol was added, followed by flash-freezing in liquid N₂ and storage at –80 °C. CbbX for X-ray crystallographic studies was purified further by gel filtration (Superdex200) in buffer A.

Enzymatic assays. All assays were performed at 25 °C. CO₂ fixation was measured in reactions containing 50 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 30 mM NaH¹⁴CO₃ (25 Bq nmol⁻¹), 3 mM RuBP, 4 mM ATP, and E.C.M., E.I. and CbbX as stated in figure legends. E.C.M. was obtained by preincubating Rubisco in the reaction mix for 20 min before the addition of RuBP. The E.I. complex was obtained by incubating Rubisco (about 100 μ M) with EDTA (4 mM final concentration) for 10 min and then adding xylulose 1,5-bisphosphate-scavenged RuBP⁴² (1 mM). Relative CbbX activities were measured similarly to the method previously described for Rca³⁶, by determining the increase in Rubisco activity during the first minute of a CO₂ fixation assay.

ATPase activity was assayed spectrophotometrically using a coupled assay that followed the oxidation of NADH³⁷. E.I. complex in the absence of free RuBP was obtained by buffer exchange using Micro Bio-Spin chromatography columns (Bio-Rad). E.C.M. used in ATPase reactions was formed by the incubation of Rubisco (about 100 μ M) in 40 mM NaHCO₃ and 10 mM MgCl₂ for 20 min.

Analysis of activation reactions by anti-Flag pull-down. Samples (50 μ l) in buffer B containing Rubisco E.I. consisting of N-terminally Flag-tagged or untagged RbcL subunits were incubated for 10 min in the presence of 2 μ M CbbX, 2 mM ATP and 1 mM RuBP at 25 °C. Subsequently, 400 μ l of buffer B was added and the samples were incubated with 20 μ l EZview Red anti-Flag M2

Affinity gel (Sigma). Rubisco consisting of co-assembled Flag-tagged and untagged RbcL subunits, contained in 500 μ l of *E. coli* soluble lysate, was subjected to pull-down as control (see above for co-expression of Flag-tagged and untagged RbcL). After incubation for 1 h at 4 °C, the beads were washed three times with 250 μ l of buffer B, eluted with SDS sample buffer and analysed by 10% SDS–PAGE and Coomassie staining.

Electron microscopy and reconstruction. CbbX (100 μ g ml⁻¹), under conditions as specified in the figure legends in buffer containing 40 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM MgCl₂, was negatively stained with 2% (w/v) uranyl acetate. Images were recorded on a Philips CM20FEG electron microscope equipped with a TEM Cam F415MP at a nominal magnification of $\times 50,000$. For three-dimensional reconstruction, images of CbbX (72 μ g ml⁻¹) were digitally recorded on a Tecnai G2 Spirit TEM with an Eagle 2,048 \times 2,048-pixel charge-coupled device camera (FEI Company). The microscope was operated under low-dose conditions at 120 keV. The images were taken at a nominal magnification of $\times 90,600$ with defocus ranging from 260 to 1,800 nm at a final sampling rate of 3.31 Å per pixel at the specimen level. A total of 1,373 particles were manually selected with the Medical Research Council program Ximdisp⁴³. The defocus and astigmatism of the images were determined with CTFFIND3 (ref. 44), and phases were corrected for effects of the contrast transfer function in SPIDER^{45,46}. Initial image processing was done with IMAGIC-5 (ref. 47). Particle images were bandpass filtered between 150 and 15 Å, normalized and centred by iteratively aligning them to their rotationally averaged sum. Initial class averages containing 10–20 images were obtained by two rounds of classification based on multivariate statistical analysis, followed by multi-reference alignment using homogenous classes as new references. A low-resolution density map was created by angular reconstitution. After Euler-angle assignment by projection matching in SPIDER, a final reconstruction was generated with imposed six-fold symmetry. Because of the preferred orientation of particles on the grid (mainly top views), the final reconstruction comprises only 245 particles. The resolution of the structure, 21 Å, is estimated by Fourier shell correlation with 0.5 correlation cut-off and loose mask. The hexameric model of CbbX was fitted with Chimera⁴⁸.

Crystallization and data collection. CbbX crystals were grown using the hanging-drop vapour-diffusion method at 18 °C by mixing 1 μ l of protein sample at 10 mg ml⁻¹ and 1 μ l of reservoir solution. Cube-shaped crystals were obtained after 4 weeks with a precipitant containing 50 mM MES-NaOH pH 6.5 and 0.4 M ammonium sulphate. For cryoprotection the crystals were transferred stepwise into mother liquor containing 1 M ammonium sulphate, 50 mM MES-NaOH pH 6.5 and 25% glycerol, and flash-frozen in liquid nitrogen.

Diffraction data were integrated and scaled with XDS⁴⁹. Pointless⁵⁰, SCALA^{51,52} and TRUNCATE⁵³ were used to convert the data to CCP4 format. The structure of CbbX was solved by SAD using crystals from selenomethionine (SeMet) labelled protein, which was expressed in M9 minimal medium supplemented with SeMet⁵⁴. Fourteen Se sites were found by direct methods using SHELXD⁵⁵. Sharp was used for the refinement of heavy-atom positions and calculation of phases⁵⁶. Density modification was performed with Resolve⁵⁷. The resulting map was readily interpretable and a structural model was manually built with Coot⁵⁸. The final model was created by using nearly isomorphous native data, performing iterative Coot model building and REFMAC5 refinement cycles^{52,59}. The final model contains two CbbX chains, six sulphate ions and 23 water molecules. In chain A, residues 147–149, 270–272 and 297–309 were disordered; in chain B, electron density for residues 1–7, 62–64 and 297–309 was not discernible. Non-glycine residues facing solvent channels without detectable side-chain density were modelled as alanines. The model has two Ramachandran outliers according to the criteria of the program PROCHECK⁶⁰.

Coordinates were aligned with Lsqkab and Lsqman⁶¹. Figures were generated with the programs PyMOL (<http://www.pymol.org>) and ESPript⁶².

39. Pitcher, D. G., Saunders, N. A. & Owen, R. J. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett. Appl. Microbiol.* **8**, 151–156 (1989).
40. Gibson, J. L., Falcone, D. L. & Tabita, F. R. Nucleotide sequence, transcriptional analysis, and expression of genes encoded within the form I CO₂ fixation operon of *Rhodobacter sphaeroides*. *J. Biol. Chem.* **266**, 14646–14653 (1991).
41. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose p-BAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
42. Edmondson, D. L., Badger, M. R. & Andrews, T. J. A kinetic characterization of slow inactivation of ribulosebiphosphate carboxylase during catalysis. *Plant Physiol.* **93**, 1376–1382 (1990).
43. Smith, J. M. Ximdisp—a visualization tool to aid structure determination from electron microscope images. *J. Struct. Biol.* **125**, 223–228 (1999).
44. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
45. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
46. Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3**, 1941–1974 (2008).

47. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
48. Pettersen, E. F. *et al.* UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
49. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
50. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
51. Evans, P. R. Scala. *CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **33**, 22–24 (1997).
52. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
53. French, G. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
54. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
55. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
56. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
57. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
59. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
60. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
61. Kleywegt, G. T. & Jones, T. A. A super position. *CCP4/ESF-EACBM Newsl. Prot. Crystallogr.* **31**, 9–14 (1994).
62. Gouet, P., Courcelle, E., Stuart, D. I. & Metz, F. ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).

Infrared diffuse interstellar bands in the Galactic Centre region

T. R. Geballe¹, F. Najarro², D. F. Figer³, B. W. Schlegelmilch⁴ & D. de la Fuente²

The spectrum of any star viewed through a sufficient quantity of diffuse interstellar material reveals a number of absorption features collectively called 'diffuse interstellar bands' (DIBs). The first DIBs were reported about 90 years ago¹, and currently well over 500 are known². None of them has been convincingly identified with any specific element or molecule, although recent studies suggest that the DIB carriers are polyatomic molecules containing carbon^{3–5}. Most of the DIBs currently known are at visible and very near-infrared wavelengths, with only two previously known at wavelengths beyond one micrometre (10,000 ångströms), the longer of which is at 1.318 micrometres (ref. 6). Here we report 13 diffuse interstellar bands in the 1.5–1.8 micrometre interval on high-extinction sightlines towards stars in the Galactic Centre. We argue that they originate almost entirely in the Galactic Centre region, a considerably warmer and harsher environment than where DIBs have been observed previously. The relative strengths of these DIBs towards the Galactic Centre and the Cygnus OB2 diffuse cloud are consistent with their strengths scaling mainly with the extinction by diffuse material.

Figure 1 shows spectra of three Galactic Centre stars, namely GCS3-2 (ref. 7), X174516 (ref. 8) and qF362 (ref. 9) in a small portion of the H band (wavelength 1.5–1.8 µm). These three stars have very different intrinsic spectra, yet they have four common absorption features in this wavelength interval. These features appear to be present in the spectra of all six Galactic Centre stars that we have observed to date, whose spectral classifications range from late B to Wolf-Rayet. A weaker fifth feature is detected in most of the stars. Their full H-band spectra show considerable diversity, as expected, with the exception of these absorptions and a number of additional ones reported below. This is evidence that the newly discovered absorption features are formed in interstellar material. Further evidence comes from their fixed wavelengths from star to star, in comparison to the wavelength shifts in the Brackett lines due to the radial velocities of the stars and in some cases their stellar winds. The strengths of the stronger features are the same to within 20% from star to star. GCS3-2 and qF362 are separated from each other on the plane of the sky by about 1 pc (we assume a Galactic Centre distance of 8 kpc), whereas X174516 is separated from each of them by about 55 pc and thus is observed on a distinctly different sightline.

Searches of atomic line databases show that none of the absorption features corresponds within 0.002 µm to atomic or ionic lines with low-lying lower energy levels. The irregular pattern of their wavelengths and their differing intensities do not lend themselves to identification as one or more simple molecules. None has been reported in the spectra of young stars embedded in dense molecular clouds. However, previously unpublished spectra of bright stars in the Cygnus OB2 cluster contain four of the five absorption features (Fig. 1). The Cygnus OB2 cluster lies behind a diffuse cloud that attenuates optical light by 5–6 mag (ref. 10) and is responsible for the numerous DIBs in the optical spectra of these stars¹¹, as well as the two previously discovered J-band DIBs^{6,12}. Finally, as discussed below, in the J-band (1.1–1.35 µm) spectrum of

qF362 (the only one of the six that we have observed in that band), the longer wavelength of the two previously discovered J-band DIBs⁶ is clearly detected at 1.318 µm; the signal-to-noise ratio at the wavelength

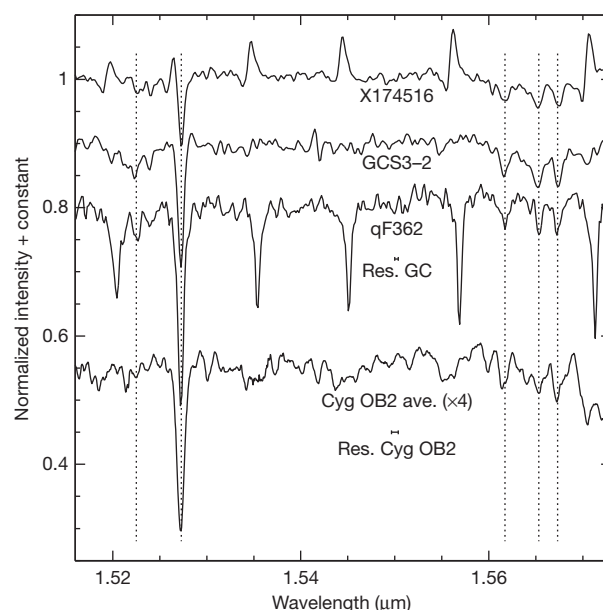


Figure 1 | Observed spectra of three hot stars in the Galactic Centre and an average spectrum of seven stars in the Cygnus OB2 association. Shown (top to bottom) are spectra of stars X174516, GCS3-2 and qF362, and the Cygnus OB2 average spectrum. The wavelengths of five newly discovered diffuse interstellar bands (DIBs) are indicated by vertical dotted lines. The prominent emission lines in the spectrum of X174516 and absorption lines in the spectrum of qF362 are Brackett series transitions of atomic hydrogen (lower-level quantum number 4 and upper-level quantum numbers 15–20), which are often present in hot and luminous stars. The strongest DIB, at 1.5273 µm, is blended with one of the Brackett lines in the spectra of X174516 and qF362, but is uncontaminated in the spectrum of GCS3-2. The observations of the Galactic Centre sources were carried out in July 2010 on Mauna Kea at the Frederick C. Gillett Gemini North Telescope using its near-infrared integral field spectrograph, NIFS. Spectra of the unreddened early A-type dwarf stars, HD 156721 and HD 174249, located outside the Galactic Centre and far from Galactic Centre sightlines and observed at similar airmasses, served as telluric standards. The prominent hydrogen absorption lines in their spectra were removed by dividing by a model spectrum of the A-star Vega. The spectra of the Galactic Centre objects were divided by these 'de-lined A-star spectra in order to remove telluric absorption lines. The continuum of each resultant spectrum was fitted with a spline function and divided by that function to produce the normalized spectra, shown here at $R \approx 4,000$ (equivalent to a wavelength resolution of 0.00038 µm at 1.54 µm; scale bar 'Res. GC'). Wavelength calibration, obtained using telluric absorption lines, is accurate to 0.0001 µm. The Cygnus OB2 average spectrum (resolution $R = 2,250$, scale bar labelled 'Res. Cyg OB2') is based on individual spectra obtained in 2002 at the Telescopio Nazionale Galileo on La Palma, which were reduced using similar techniques as described below.

¹Gemini Observatory, 670 N. A'ohoku Place, Hilo, Hawaii 96720, USA. ²Centro de Astrobiología (CSIC-INTA), Ctra. Torrejón a Ajalvir km 4, 28850 Torrejón de Ardoz, Spain. ³Center for Detectors, Rochester Institute of Technology, Rochester, New York 14623, USA. ⁴Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA.

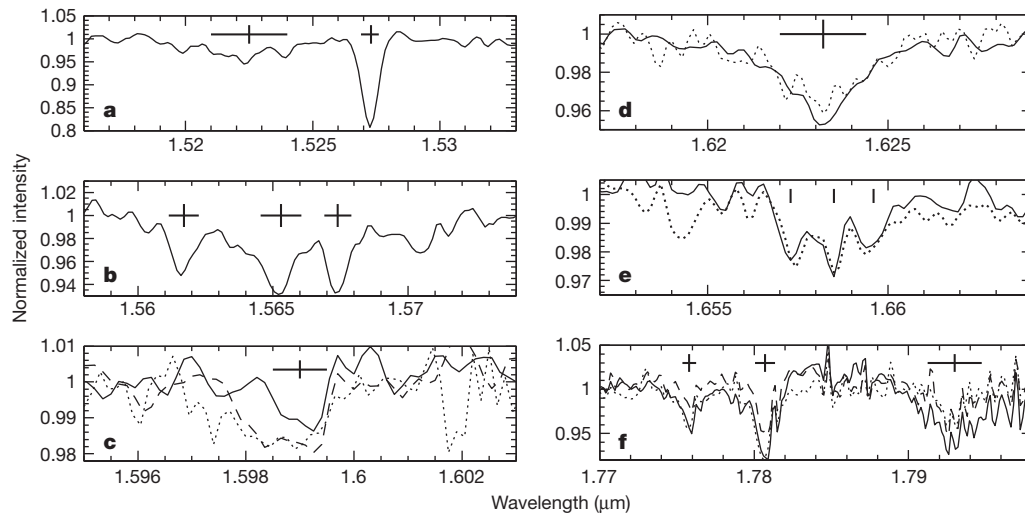


Figure 2 | Spectra of the newly discovered DIBs. Vertical lines in **a–f** indicate peak wavelengths, and horizontal lines the observed full-widths at half-maximum. Spectra are shown for three stars: GCS3-2 (continuous lines), X174516 (dashed lines) and qF362 (dotted lines). The DIBs are at the following wavelengths: 1.5225, 1.5273, 1.5617, 1.5653, 1.5673, 1.5990, 1.6232, 1.6573, 1.6585, 1.6596, 1.7758, 1.7807 and 1.7930 μm . Most are considerably weaker than the four strongest ones (see Fig. 1 and **a** and **b** here). In **c–f**, the spectrum of more than one star is shown. As the spectra of the individual stars are diverse, we cannot rule out the presence of contaminating stellar lines in some of these spectral regions. The spectrum of GCS3-2 is the least contaminated by stellar

lines; thus its spectrum has been weighted most strongly by us in deciding which absorption features are DIBs. The H-band DIBs have a wide range of spectral widths. The optical depth of the strongest DIB at 1.5273 μm is at least twice that of any of the other H-band DIBs. Panels **b**, **e**, **f** contain absorption triplets, centred near 1.565 μm , 1.658 μm and 1.784 μm , respectively. The triplet near 1.658 μm (in **e**) appears to be partially blended and might be regarded as one DIB. However, as discussed in the text, the large velocity dispersion of the diffuse gas in which we believe the bulk of the absorptions arise may be mostly responsible for this blending.

of the shorter wavelength J-band DIB is too low to yield a high-confidence detection.

We conclude, on the basis of all the above evidence, that the newly discovered features are members of the family of DIBs. They are the longest-wavelength DIBs reported to date. Our longer-wavelength K-band (2.0–2.4 μm) spectra of these stars do not reveal additional DIBs, to optical depth limits of ~ 0.02 or better over most of that band. Figure 2 shows spectra of the complete set of newly discovered H-band DIBs, and Table 1 lists their properties.

Where along the 8-kpc-long sightline to the Galactic Centre are the carriers of the H-band DIBs located? The dust along that sightline produces approximately 30 visual magnitudes of extinction⁹, rendering the Galactic Centre unobservable at optical wavelengths. Approximately one-third of the extinction occurs in dense molecular clouds, presumably located in intervening spiral arms, whereas the remaining

two-thirds arises in diffuse clouds¹³. The newly discovered DIBs are thus associated with about 20 visual magnitudes of extinction. It is not surprising then that they are considerably stronger towards the Galactic Centre than towards Cygnus OB2.

There is strong evidence that the diffuse interstellar gas on the sightline to the Galactic Centre is found, almost in its entirety, very close to the centre, in a region of radius ~ 200 pc known as the Central Molecular Zone (CMZ). Spectroscopy of the molecular ion H_3^+ towards Galactic Centre sources, including GCS3-2, has shown that the CMZ contains a vast quantity of diffuse gas^{14,15}. Because the temperature of the gas in the CMZ, which can be measured using H_3^+ , is much higher than in diffuse clouds in the Galactic plane, and because of the high velocity dispersion of the gas in the CMZ ($\sim 150 \text{ km s}^{-1}$ in the direction of GCS3-2; ref. 14), high-resolution spectroscopy of H_3^+ can be used to clearly distinguish between diffuse material inside and outside the CMZ. There is no evidence on these sightlines for a significant fraction of the diffuse cloud material being located outside the CMZ. If, as we suspect, the H-band DIBs largely originate in the CMZ's diffuse gas, their widths are probably Doppler-broadened similarly to H_3^+ . That broadening, corresponding to $\sim 0.0008 \mu\text{m}$ in the H band, is greater than the resolution of the spectrum ($0.0004 \mu\text{m}$). Thus some of the newly discovered DIBs may have intrinsic spectral widths that are much less than they appear towards the Galactic Centre, perhaps only a few ångströms. This is not unusual, as many DIBs at visual wavelengths have widths of $\sim 1 \text{ Å}$ (ref. 16). However, as the precise velocity distribution of the material producing these DIBs is not known, the laboratory wavelengths of the new bands currently should be considered accurate to only $\pm 0.0010 \mu\text{m}$ ($\pm 10 \text{ Å}$).

Spectroscopy of Galactic Centre sources is much more difficult in the J band than in the H band, because the attenuation by dust is nearly an order of magnitude greater in the former. Nevertheless, we have secured a J-band spectrum of qF362, in which the previously discovered 1.318 μm DIB is prominently present. Figure 3 shows the portion of the spectrum containing this DIB. The strengths of the 1.318- μm and 1.527- μm DIBs in the Galactic Centre sources are both about four times greater than they are in the spectra of stars in Cygnus OB2. If the strengths of these DIBs scale roughly with extinction¹⁶, they imply that

Table 1 | Properties of DIBs measured towards GCS3-2

Wavelength* (μm)	FWHM† (Å)	$\tau_{\text{obs}}^\ddagger$ (max.)	W_λ § (Å)	No. of detections
1.5225	30 ± 10	0.04 ± 0.01	1.6 ± 0.4	4
1.5273	6 ± 1	0.19 ± 0.01	1.5 ± 0.1	6
1.5617	10 ± 2	0.05 ± 0.01		5*
1.5653	15 ± 4	0.07 ± 0.01	$2.8 \pm 0.3^\ddagger$	6
1.5673	9 ± 2	0.07 ± 0.01		6
1.5990	9 ± 2	0.015 ± 0.007	0.15 ± 0.05	4
1.6232	24 ± 3	0.045 ± 0.005	1.3 ± 0.3	6
1.6573	ND	0.022 ± 0.006		4
1.6585	ND	0.028 ± 0.006	$0.56 \pm 0.12^\ddagger$	6
1.6596	ND	0.018 ± 0.006		5
1.7758	8 ± 2	0.05 ± 0.01	0.6 ± 0.2	6
1.7807	12 ± 3	0.08 ± 0.01	1.0 ± 0.3	6
1.7930	35 ± 15	0.06 ± 0.02	1.5 ± 0.5	6

FWHM, full-width at half-maximum; τ_{obs} , observed optical depth; W_λ , equivalent width; ND, not determined.

* Uncertainty in wavelength is $0.0010 \mu\text{m}$; values obtained assuming peak absorption occurs at $v(\text{LSR}) = 0 \text{ km s}^{-1}$, where $v(\text{LSR})$ is the radial velocity relative to the Local Standard of Rest.

† Visual estimates; uncertainties depend on strength of DIB and signal-to-noise ratio in the wavelength region of the DIB.

‡ Deconvolved, assuming Gaussian DIB profile and Gaussian instrumental profile.

§ Determined by numerical integration.

|| Towards Galactic Centre sources, out of six sightlines observed (three shown here, three not shown).

* Includes W_λ of DIBs at adjacent wavelengths.

† Probably present in the sixth star, but low signal-to-noise ratio of spectrum, and strong and broad emission lines, prohibit clear detection.

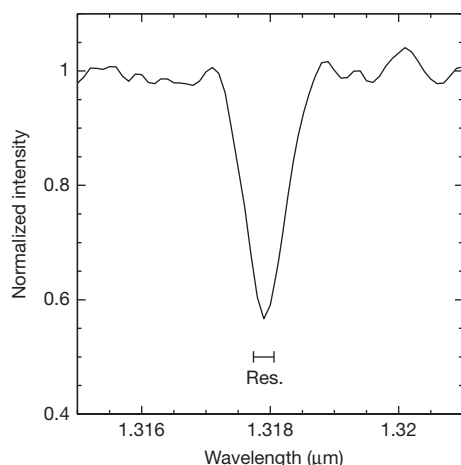


Figure 3 | Profile of the 1.318- μm diffuse interstellar band towards qF362. The equivalent width of the feature is 3.13 \AA , with an uncertainty of $\pm 0.10 \text{ \AA}$, compared to $0.86 \pm 0.06 \text{ \AA}$ in the Cygnus OB2 cluster star BD +40°42203, which suffers a visual extinction of 5.97 mag (ref. 20). Scaling extinction with the equivalent width of the 1.318 μm feature results in an estimated diffuse cloud visual extinction towards qF362 of $21.7 \pm 1.7 \text{ mag}$, which is consistent with a previously estimate of diffuse cloud extinction towards the Galactic Centre¹². The resolution of the spectrum is $R \approx 4,000$ (equivalent to a wavelength resolution of $0.00033 \mu\text{m}$ at $1.32 \mu\text{m}$; scale bar 'Res.').

there are $\sim 20 \text{ mag}$ of visual extinction in diffuse clouds towards the Galactic Centre, which is consistent with previous estimates¹³.

The diffuse interstellar medium in the Galactic Centre is a considerably harsher environment than the diffuse clouds where DIBs have been previously observed. The gas temperatures in the centre are 200–300 K (ref. 13), compared to 30–100 K in Galactic diffuse clouds¹⁸, and the cosmic ray ionization rate is an order of magnitude higher^{14,17}. The finding that the strengths of the J-band and H-band DIBs in Cyg OB2 and the Galactic Centre are in rough proportion to the extinction suggests that the carriers of these bands survive equally well in both environments.

Interstellar extinction is mainly caused by silicate-based dust particles¹⁸, and thus it is not obvious that the strengths of the 1.318- μm and 1.527- μm DIBs should depend only on the extinction in diffuse clouds if, as suspected, their carriers are carbonaceous. Indeed, in the Galactic plane the strengths of many visual wavelength DIBs do not correlate well with extinction¹⁹ or with one another from source to source^{4,19}. Owing to the high extinction towards the Galactic Centre, no information on the strengths of visual DIBs in that direction is available, and thus comparisons of them with the new H-band DIBs cannot be made.

In contrast to the H-band DIBs, the strength of the 3.4- μm interstellar hydrocarbon absorption band is 2–3 times higher towards the Galactic Centre than one would predict by scaling its strength in Galactic diffuse clouds by the ratio of extinctions²⁰. The different behaviours of two families of probable carbonaceous particles—the carriers of the H-band DIBs and the carriers of the 3.4- μm feature—indicate that their abundances respond differently to the different physical conditions inside and outside the Galactic Centre, and suggest that the carriers are not closely coupled by interstellar chemistry. A similar conclusion has been drawn for a limited number of visual wavelength DIBs observed in several other sightlines²¹. Differences in the interstellar carbon abundance in diffuse clouds in the Galactic

plane and in the Galactic Centre could also affect the relative strengths of the bands in those locations, but little information on the carbon abundance in the Galactic Centre is available.

Received 14 July; accepted 31 August 2011.

Published online 2 November 2011.

- Heger, M. L. The spectra of certain class B stars in the regions 5630–6680 \AA and 3280–3380 \AA . *Lick Obs. Bull.* **10**, 146–147 (1922).
- DIB database. (<http://dibdata.org>) (July 2011).
- Cox, N. L. J. in *PAHs and the Universe: A Symposium to Celebrate the 25th Anniversary of the PAH Hypothesis* (eds Joblin, C. & Tielens, A. G. G. M.) 349–354 (EAS Publication Series Vol. 46, European Astronomical Society, 2011).
- Snow, T. P. & McCall, B. J. Diffuse atomic and molecular clouds. *Annu. Rev. Astron. Astrophys.* **44**, 367–414 (2006).
- Oka, T. & McCall, B. J. Disclosing identities in diffuse interstellar bands. *Science* **331**, 293–294 (2011).
- Joblin, C., Maillard, J. P., D'Hendecourt, L. & Léger, A. Detection of diffuse interstellar bands in the infrared. *Nature* **346**, 729–731 (1990).
- Nagata, T., Woodward, C. E., Shure, M., Pipher, J. L. & Okuda, H. AFGL 2004: an infrared quintuplet near the Galactic Center. *Astrophys. J.* **351**, 83–88 (1990).
- Muno, M. P. et al. Isolated massive supergiants near the Galactic Center. *Astrophys. J.* **638**, 183–190 (2006).
- Figer, D. F., McLean, I. S. & Morris, M. Massive stars in the Quintuplet Cluster. *Astrophys. J.* **514**, 202–220 (1999).
- Humphreys, R. M. Studies of luminous stars in nearby galaxies. I. Supergiants and O stars in the Milky Way. *Astrophys. J.* **38**, 309–350 (1978).
- Bromage, G. E. & Nandy, K. Observations of diffuse interstellar features in the spectra of dust-embedded and field stars. *Astron. Astrophys.* **26**, 17–32 (1973).
- Rawlings, M., Adamson, A., McCall, B. & Kerr, T. A high-resolution study of the near-infrared diffuse interstellar bands. Presented at IAU Symp. No. 280 (<http://adsabs.harvard.edu/abs/2011IAUS..280P.311R>) (2011).
- Whittet, D. C. B. et al. Infrared spectroscopy of dust in the diffuse interstellar medium toward Cygnus OB2 no. 12. *Astrophys. J.* **490**, 729–734 (1997).
- Oka, T., Geballe, T. R., Goto, M., Usuda, T. & McCall, B. J. Hot and diffuse clouds near the Galactic Center probed by metastable H_3^+ . *Astrophys. J.* **632**, 882–893 (2005).
- Goto, M. et al. Absorption line survey of H_3^+ toward the Galactic Center sources. II. Eight infrared sources within 30 pc of the Galactic Center. *Astrophys. J.* **688**, 306–319 (2008).
- Hobbs, L. M. et al. Studies of the diffuse interstellar bands. III. HD 183143. *Astrophys. J.* **705**, 32–45 (2009).
- Indriolo, N., Geballe, T. R., Takeshi, O. & McCall, B. J. H_3^+ in diffuse interstellar clouds: a tracer for the cosmic-ray ionization rate. *Astrophys. J.* **671**, 1736–1747 (2007).
- Draine, B. T. *Physics of the Interstellar and Intergalactic Medium* (Princeton Univ. Press, 2011).
- Herbig, G. H. The diffuse interstellar bands. *Annu. Rev. Astron. Astrophys.* **33**, 19–73 (1995).
- Pendleton, Y. J., Sandford, S. A., Allamandola, L. J., Tielens, A. G. G. M. & Sellgren, K. Near-infrared absorption spectroscopy of interstellar hydrocarbon grains. *Astrophys. J.* **437**, 683–696 (1994).
- Rawlings, M. G., Adamson, A. J. & Whittet, D. C. B. Infrared and visual interstellar absorption features toward heavily reddened field stars. *Mon. Not. R. Astron. Soc.* **341**, 1121–1140 (2003).

Acknowledgements This Letter is based on observations obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (US), the Science and Technology Facilities Council (UK), the National Research Council (Canada), CONICYT (Chile), the Australian Research Council (Australia), Ministério da Ciência e Tecnologia (Brazil) and Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina). This work was supported by the Spanish Ministerio de Ciencia e Innovación. We thank A. Lenorzer for reductions of the H-band spectra of the Cygnus OB2 stars, and B. J. McCall and T. Oka for reviewing a preliminary version of the manuscript.

Author Contributions T.R.G. and F.N. wrote the observing proposal. T.R.G. obtained the data. B.W.S. and T.R.G. reduced the data. F.N. and T.R.G. recognized the spectral features as DIBs. T.R.G., F.N., D.F.F., D.F. and B.W.S. discussed the results. T.R.G., F.N. and D.F.F. wrote the Letter.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.R.G. (tgeballe@gemini.edu).

Two types of luminescence blinking revealed by spectroelectrochemistry of single quantum dots

Christophe Galland^{1,2}, Yagnaseni Ghosh³, Andrea Steinbrück³, Milan Sykora¹, Jennifer A. Hollingsworth³, Victor I. Klimov^{1,2} & Han Htoon^{1,2,3}

Photoluminescence blinking—random switching between states of high (ON) and low (OFF) emissivities—is a universal property of molecular emitters found in dyes¹, polymers², biological molecules³ and artificial nanostructures such as nanocrystal quantum dots, carbon nanotubes and nanowires^{4–6}. For the past 15 years, colloidal nanocrystals have been used as a model system to study this phenomenon^{5,6}. The occurrence of OFF periods in nanocrystal emission has been commonly attributed to the presence of an additional charge⁷, which leads to photoluminescence quenching by non-radiative recombination (the Auger mechanism)⁸. However, this ‘charging’ model was recently challenged in several reports^{9,10}. Here we report time-resolved photoluminescence studies of individual nanocrystal quantum dots performed while electrochemically controlling the degree of their charging, with the goal of clarifying the role of charging in blinking. We find that two distinct types of blinking are possible: conventional (A-type) blinking due to charging and discharging of the nanocrystal core, in which lower photoluminescence intensities correlate with shorter photoluminescence lifetimes; and a second sort (B-type), in which large changes in the emission intensity are not accompanied by significant changes in emission dynamics. We attribute B-type blinking to charge fluctuations in the electron-accepting surface sites. When unoccupied, these sites intercept ‘hot’ electrons before they relax into emitting core states. Both blinking mechanisms can be electrochemically controlled and completely suppressed by application of an appropriate potential.

In the conventional blinking model (Fig. 1a), ON and OFF periods correspond to a neutral nanocrystal and a charged nanocrystal, respectively, and photo-assisted charging/discharging causes random switching between these two states. The dynamics of the bright state is dominated by radiative recombination of the neutral exciton, X^0 (Fig. 1b), which is characterized by a long, mono-exponential decay (15–30 ns in CdSe nanocrystals^{11–13}). For a charged exciton (trion), X^- , three-particle Auger recombination opens a fast, non-radiative channel, resulting in a shorter lifetime (a few nanoseconds or less) and, consequently, a reduced photoluminescence quantum yield. As illustrated in Figs 1c, d, this model predicts correlated fluctuations of the photoluminescence intensity and lifetime (referred to here as A-type blinking) that have indeed been observed experimentally^{11,12,14}. Photocharging can lead to ‘binary’ switching between the ON and the OFF states (Fig. 1c) when the timescale of charge fluctuations is longer than the experimental binning time (typically at least tens of milliseconds). As we discuss below, the same discrete charging process can also produce quasi-continuous photoluminescence fluctuations, referred to as ‘flickering’ (Fig. 1d). In this case, the data within each bin represents an average over the neutral and charged nanocrystal states, which results in photoluminescence intensities and lifetimes that vary continuously according to the relative times spent by the nanocrystal in each charge state.

A convenient tool for the analysis of correlations between photoluminescence intensities and lifetimes is a fluorescence lifetime–intensity

distribution (FLID) representation. In this representation, the probability of occupying a given state in the two-dimensional lifetime–intensity space is shown by false colour, which changes from blue to red as the probability increases. As illustrated in the insets of Figs 1c, d, the use of FLIDs allows ready identification of different types of charge state as well as different types of blinking behaviour (for example binary blinking versus flickering).

In our work, to verify the validity of the charging model of photoluminescence intermittency, we combine single-nanocrystal spectroscopy with an electrochemical approach for controlling the extent of nanocrystal charging^{15–17}. Specifically, we conduct single-nanocrystal, time-tagged, time-resolved, single-photon counting studies of samples incorporated into a three-electrode electrochemical cell (Fig. 2a). We investigate core–shell CdSe/CdS nanocrystals synthesized according to ref. 18. In the case of exceptionally thick, 16–19-monolayer shells, they show nearly complete suppression of blinking¹⁸. Here we use nanocrystals with intermediate shell thicknesses (7–9 monolayers) that show typical blinking behaviour but have quantum yields during

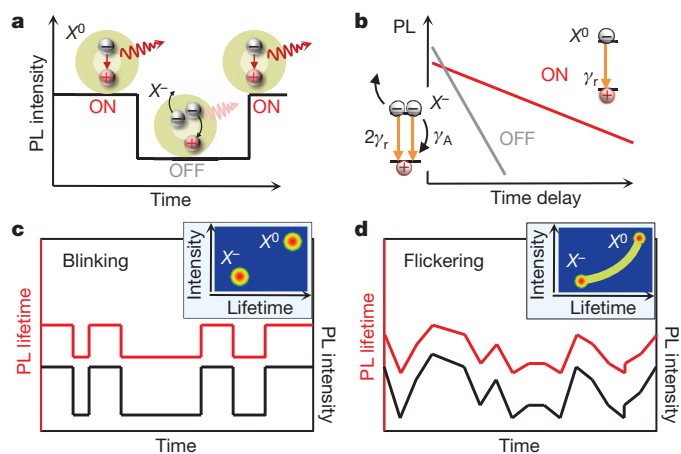


Figure 1 | Conventional charging model: A-type blinking and flickering.

a, In the conventional photoluminescence (PL) blinking model, ON and OFF periods correspond to a neutral nanocrystal (X^0) and a charged nanocrystal (X^-), respectively. **b**, Schematic photoluminescence decay of the ON and the OFF states on a logarithmic scale. The dynamics of the ON state is dominated by the radiative rate γ_r . In the charged state, the increase in the number of recombination pathways leads to a higher radiative rate, $2\gamma_r$, responsible for the higher emission intensity at short delays. Simultaneously, the onset of three-particle Auger recombination with the rate $\gamma_A \gg \gamma_r$ opens a new, non-radiative, channel, leading to faster photoluminescence decay and reduced photoluminescence quantum yield. **c**, When the timescale of charging and discharging is longer than the experimental binning time, binary blinking is observed. **d**, For fluctuations much faster than the bin size, a continuous distribution of intensities and lifetimes is obtained, often referred to as flickering. The insets in **c** and **d** show corresponding schematic fluorescence lifetime–intensity distributions (FLIDs).

¹Chemistry Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ²Center for Advanced Solar Photophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ³Materials Physics & Applications: Center for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

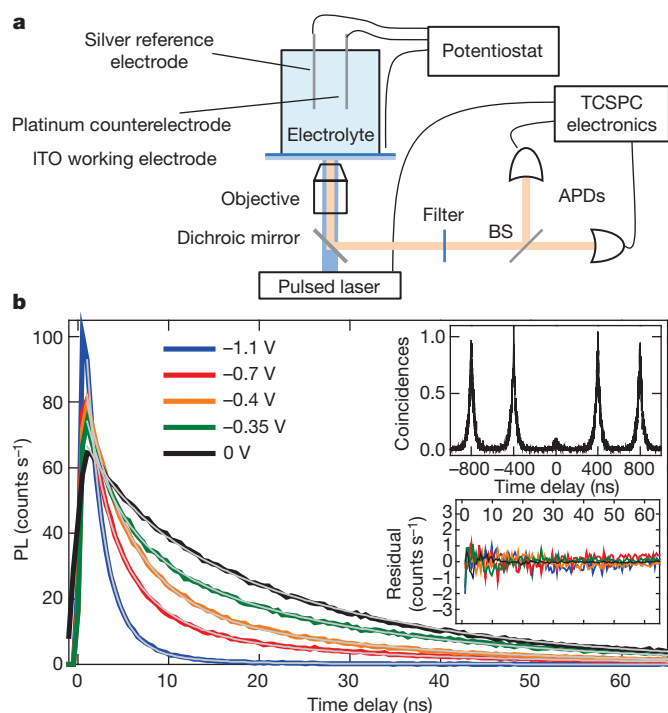


Figure 2 | Experimental set-up and electrochemical charging of an individual nanocrystal. **a**, Set-up of a single-nanocrystal spectroelectrochemical experiment. APD, avalanche photodiode; BS, 50/50 beam splitter; ITO, indium tin oxide; TCSPC, time-correlated single-photon counting. **b**, Series of photoluminescence decays for a single nanocrystal for increasingly negative applied potentials. The thin grey lines show the best global triple-exponential fits with the shared time constants, yielding the lifetimes $\tau^d = 2$ ns, $\tau^s = 5$ ns and $\tau^n = 24$ ns. Top inset: the second-order photoluminescence intensity correlation function measured for this nanocrystal indicates that $g_2(0) = 0.08$. Bottom inset: residuals of the global fit indicate very high fidelity of the fitting procedure, with deviations within the noise level and below 1% of the maximum photoluminescence signal.

OFF periods that are considerably higher than those of standard nanocrystals. This allows us to investigate in detail the properties of the OFF state and the effect of controlled charging on its emissivity and dynamics. All changes in photoluminescence intensity and dynamics induced by the applied potential are reversible, indicating no permanent chemical or photochemical modifications of the nanocrystals and suggesting that such changes are due to controlled charging/discharging. To verify that the observed photoluminescence originates from a single nanocrystal, we measure the second-order intensity correlation function¹⁹, g_2 , and ensure that $g_2(0) < 0.5$ (Methods Summary; Fig. 2b, inset). All experiments are performed under ambient conditions at room temperature. All potentials are reported with respect to a silver wire quasi-reference (see Methods for further experimental details).

Electrochemical control of emission intensity from individual nanocrystals has been demonstrated previously¹⁷. In this study, we analyse the effect of charging on both photoluminescence intensity and photoluminescence dynamics. Figure 2b shows photoluminescence time transients recorded for a single nanocrystal under increasing negative potential, V , which corresponds to electron injection. The photoluminescence decay becomes progressively faster as V becomes more negative. All decays can be fitted globally to a triple-exponential function (Fig. 2b, grey lines). The high fidelity of the fit (see residuals in the lower inset) suggests that only three distinct emitting states are involved, each having a well-defined photoluminescence lifetime: $\tau^d = 2$ ns, $\tau^s = 5$ ns and $\tau^n = 24$ ns. As V is decreased from 0 to -0.7 V, the weight of the τ^s component gradually increases relative to that of the τ^n component. At more-negative values of V , the component with the fastest decay (corresponding to τ^d) emerges and becomes dominant at

-1.1 V. We assign the lifetimes τ^n , τ^s and τ^d to three distinct states, respectively neutral excitons (X^0) and singly (X^-) and doubly (X^{2-}) charged excitons. Owing to increased rates of radiative decay, singly charged excitons (negative trions) show increased signal at short times relative to neutral excitons (Fig. 2b); however, the time-integrated photoluminescence signal is reduced because of Auger recombination. The observed quickening of photoluminescence decay on charging is due to enhancement in both the radiative decay rate and the non-radiative decay rate.

To confirm the above assignments and investigate the relationship between charging and blinking, we analyse the correlations in the temporal variations of photoluminescence decay time and intensity. In Fig. 3, we plot photoluminescence intensity and average lifetime trajectories (calculated for a 50-ms bin size; Supplementary Information) along with corresponding FLIDs for the nanocrystal shown in Fig. 2. To illustrate the variability in blinking behaviours, we present the data collected for this nanocrystal on two different days. These data, representing examples of binary ON–OFF switching (Fig. 3a) and flickering (Fig. 3b), indicate a strong correlation between the photoluminescence intensity and the photoluminescence lifetime during the fluctuations, in agreement with the conventional charging model. We call this A-type blinking.

At 0 V (Fig. 3a, middle), the nanocrystal shows binary blinking between the neutral state (X^0) and the singly charged state (X^-). The

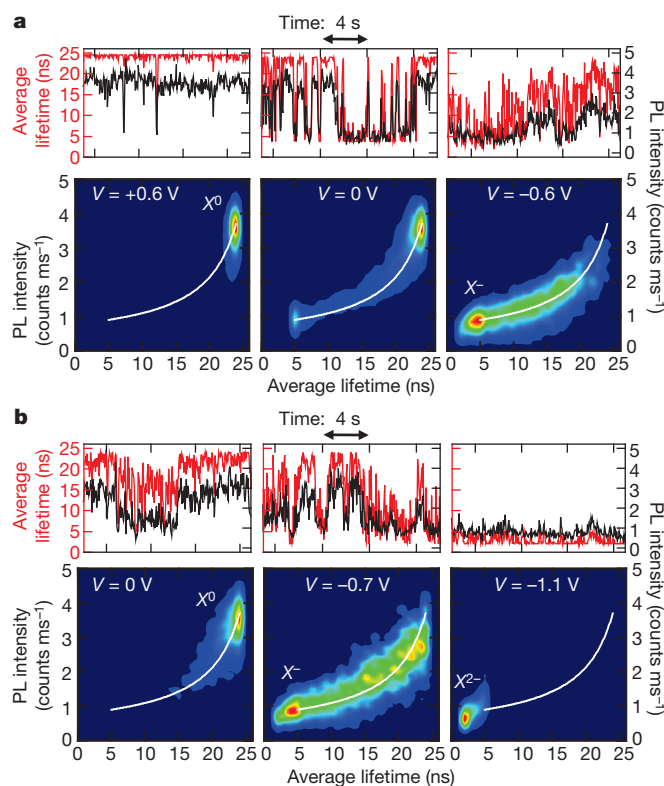


Figure 3 | Correlated photoluminescence intensity and lifetime fluctuations: A-type blinking and flickering. **a**, Photoluminescence intensities (black lines) and average lifetimes (red lines), and corresponding FLIDs, for the nanocrystal shown in Fig. 2 at three different potentials. Binary blinking seen at $V = 0$ V is largely suppressed at $V = +0.6$ V, whereas electron injection is achieved at $V = -0.6$ V. In the FLID colour scale, red corresponds to the most frequently occurring lifetime–intensity pair, and probabilities less than 1% of this maximum are indicated by dark blue. A linear scaling from blue to red is used between these extremes. **b**, Data from the same nanocrystal, acquired on a different day, display continuous photoluminescence intensity and lifetime fluctuations, typical of flickering. At $V = -1.1$ V, we observe emission from a doubly charged exciton, X^{2-} . All data were analysed with a bin size of 50 ms. Full time trajectories for **a** and **b** are shown in Supplementary Fig. 1.

average photoluminescence lifetime of X^0 (the ON state) is $\tau^n \approx 24$ ns, which corresponds to a radiative lifetime of ~ 60 ns (Supplementary Information) and is in agreement with previous ensemble studies of this type of nanocrystal¹³. Application of a positive potential, $V = +0.6$ V (Fig. 3a, left), drastically suppresses charge fluctuations and results in almost non-blinking emission from the neutral exciton (see corresponding FLID and Supplementary Figs 1 and 2; another example is shown in Supplementary Fig. 12). At a negative potential, $V = -0.6$ V (Fig. 3a, right), the peak of the photoluminescence distribution shifts to the lower-emissivity state, X^- , characterized by a lifetime of ~ 5 ns. Assuming a 'statistical' scaling of recombination rates with the number of charges (Supplementary Information and ref. 20), we deduce the Auger lifetime for X^- to be ~ 3.5 ns. This is much shorter than the radiative lifetime of X^- (~ 30 ns; Supplementary Information), which explains the relatively low photoluminescence quantum yield of the negative trion. The existence of fluctuations between X^0 and X^- is indicated by a well-resolved trace in the FLID connecting the two states. We simulate the FLID data assuming that the photoluminescence intensity during a given time bin is determined by the relative times spent by the nanocrystal in the states X^- and X^0 (Supplementary Information). A very good agreement, without any adjustable parameters, between the simulated trace (Fig. 3a, white lines) and the measured FLID provides strong support for both the assignment of emitting states and the model used in the analysis.

We note that the same nanocrystal measured on a different day (Fig. 3b) shows a more continuous distribution of photoluminescence intensities and lifetimes, typically referred to as flickering. This change in the blinking behaviour probably occurs as a result of the shortening of time spent by the nanocrystal in a given charge state, which leads to fast switching between X^0 and X^- within the bin time used in the measurements. Photoluminescence from X^- becomes dominant at $V = -0.7$ V (Fig. 3b, middle FLID). By applying a more negative potential, $V = -1.1$ V, we detect a new state with lifetime $\tau^d \approx 2$ ns, associated with the formation of a doubly charged exciton, X^{2-} , with an Auger lifetime of ~ 1.2 ns. Judging from the FLID at this potential (Fig. 3b, right), fluctuations occur also between the states X^{2-} and X^- .

Figure 4a shows data from a different nanocrystal, which has distinct blinking behaviour that we refer to as B-type blinking. Specifically, at $V = 0$ V (Supplementary Fig. 3) and $V = +0.8$ V (Fig. 4a, left), we observe periods of low photoluminescence intensity that are not accompanied by significant shortening of photoluminescence lifetimes. In fact, the photoluminescence time constant measured for the B-type OFF state is identical to that of the ON state (X^0). These B-type blinking events were observed in 20 of the 23 dots we studied (Supplementary Table 1) and usually coexisted with A-type fluctuations (Supplementary Fig. 3). Notably, at $V = -1$ V there is complete suppression of blinking but the long photoluminescence lifetime (~ 26 ns) typical of a neutral exciton is preserved. This suppression could be achieved in the majority of the nanocrystals with B-type blinking; however, the potential required to obtain the suppression varied widely from dot to dot (from -0.6 to -1.4 V; Supplementary Table 1). For some nanocrystals, the elimination of B-type blinking occurred simultaneously with the onset of A-type fluctuations between X^0 and X^- (see below). At a more negative potential ($V = -1.2$ V; Fig. 4a, right), we observe clear signatures of electron injection into the nanocrystal. The photoluminescence decay becomes bi-exponential, with an increasing contribution from the negative trion, which in this quantum dot has a lifetime of ~ 6 ns (Supplementary Fig. 4). In this case, switching between X^0 and X^- occurs on a much shorter timescale than the bin time, which gives rise to a narrow photoluminescence lifetime-intensity distribution. As with the data in Fig. 3, we can closely reproduce this pattern using the charging model (simulated white lines in FLID).

To explain B-type blinking, we invoke the activation and deactivation of non-radiative recombination centres (denoted R) that efficiently capture 'hot' electrons before they relax into the lowest-energy emitting

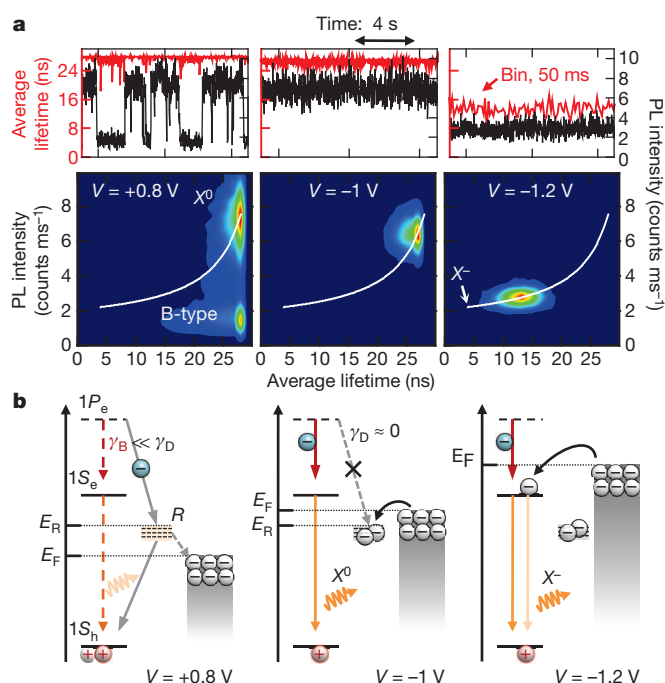


Figure 4 | Photoluminescence intensity fluctuations without lifetime changes: B-type blinking. **a**, Photoluminescence intensities (black lines) and average lifetimes (red lines), and corresponding FLIDs, for a nanocrystal showing the B-type OFF state; analysis done with a 10-ms bin. Full time trajectories are shown in Supplementary Fig. 6. **b**, The model of B-type blinking. The B-type OFF state is due to the activation of recombination centres (R) that capture hot electrons at a rate, γ_D , that is higher than the intraband relaxation rate, γ_B (the ground and the excited electron states are shown as $1S_e$ and $1P_e$, respectively; $1S_h$ is the band-edge hole state). The position of the Fermi level, E_F , relative to the trap energy, E_R , is determined by the electrochemical potential and controls the occupancy of the surface trap R . This, in turn, allows for electrochemical control of B-type blinking.

state (Fig. 4b). Such processes of hot-electron trapping have been recently observed for both nanocrystals in solutions²¹ and surface-dispersed particles^{22,23}. In this picture, photoluminescence dynamics during the OFF periods should be similar to that of a neutral exciton whereas the emission intensity will be reduced according to the ratio between the rates of intraband relaxation, γ_B , and hot-electron capture by the recombination centre, γ_D . Because the frequency of B-type blinking events is controlled by the electrochemical potential, the activation and deactivation of the bypass channel are probably associated with emptying and, respectively, filling of the corresponding surface trap state. For a positive potential ($V = +0.8$ V; Fig. 4a, left), the Fermi level decreases in energy, which increases the relative time spent by the trap in the unoccupied (that is, active) state and leads to increased occurrence of B-type OFF events (Fig. 4b, left). The trapped electron can recombine non-radiatively with a valence-band hole before the next photoexcitation event, leaving behind a neutral dot. Occasionally, photon absorption occurs before reneutralization of the dot, resulting in a positive trion, X^+ ; Auger decay of X^+ could explain observations of shorter photoluminescence lifetimes within the B-type OFF periods illustrated in Supplementary Fig. 5.

For an increasingly negative potential, the Fermi level increases in energy and eventually a regime is reached where the trap states become populated and $\gamma_D \rightarrow 0$ owing to Coulomb blockade ($V = -1$ V; Fig. 4a, b, middle). In this case, B-type blinking is completely suppressed. Application of an even more negative potential leads to charging of the nanocrystal core with an extra electron and emission from negative trions ($V = -1.2$ V; Fig. 4a, b, right).

Blinking suppression due to filling of electron-accepting trap sites is consistent with previous observations that electron-donating thiols

enhance ensemble photoluminescence emission²⁴ and reduce blinking²⁵. Similar phenomena were observed for other electron-donating molecules²⁶ as well as n-doped substrates²⁷. These observations of a significant effect of surface species on photoluminescence intensity and intermittency imply that the trap sites responsible for B-type blinking are probably of surface origin. Recent ultrafast studies of carrier surface trapping in ensembles of CdSe nanocrystals also suggest that this process is directly relevant to the problem of nanocrystal blinking²⁸. Finally, our model of B-type hot-electron surface traps provides an explanation of previously reported properties of the nanocrystal OFF state, such as low emission quantum yields⁹ and the lack of a systematic size dependence of photoluminescence lifetimes¹⁰, that could not be explained by the traditional charging model (Supplementary Information, section IV).

The distinct nature of the processes responsible for A- and B-type blinking is evident from the effect of increasing shell thickness on photoluminescence intermittency. Specifically, we observe that as the outer shell gets thicker, the B-type type blinking events become less frequent until they are completely eliminated for shells with 15 or more CdS monolayers. By contrast, the A-type blinking can still be observed even in the case of the extremely thick 19-monolayer shells. The analysis of photoluminescence intermittency in more than twenty nanocrystals with 15-monolayer shells (Supplementary Table 2 and Supplementary Figs 7 and 8) indicates that ~70% of these dots are non-blinking and that the rest have A-type blinking behaviour; none of the nanocrystals showed any detectable B-type blinking. By contrast, B-type blinking is clearly the dominant behaviour in nanocrystals with 7–9-monolayer shells (Supplementary Table 1). The fact that B-type blinking is quickly suppressed as shell thickness increases is consistent with the proposed mechanism of hot-electron tunnelling outside the nanocrystal, because this process is expected to be extremely (in fact exponentially) sensitive to the thickness of the tunnelling barrier.

The studies of statistics of ON and OFF times also indicate a clear distinction between the A-type and B-type blinking mechanisms. In

Fig. 5a, we show a nanocrystal with B-type blinking at -0.8 V, which switches to A-type blinking at -1 V. Remarkably, whereas the B-type ON and OFF times both follow a power-law distribution over almost three decades, the distributions of ON and OFF times in the A-type blinking regime are quasi-exponential with a cut-off time of ~ 70 ms (Fig. 5b). This electrochemically controlled switching between different blinking regimes in the same nanocrystal is another strong indication that the difference between A-type and B-type blinking is linked to the distinct nature of the underlying physical mechanisms but not to dot-to-dot variations. Furthermore, the fact that the cut-off time measured in the case of A-type blinking is close to a typical bin size used in the measurements suggests that relatively small changes in the time-scale of charge fluctuations can result in switching between binary blinking and flickering as seen, for example, in Fig. 3.

METHODS SUMMARY

We used a home-built electrochemical cell with a three-electrode configuration. The nanocrystals were directly deposited onto an ITO-coated transparent working electrode from a very dilute hexane or water solution. As a counterelectrode, we used platinum gauze attached to a platinum wire. All potentials reported in the main text are measured relative to a silver wire quasi-reference. The electrochemical experiments were performed using several combinations of solvents (acetonitrile and propylene carbonate) and supporting electrolytes (all concentrations, 0.1 M): tetrabutylammonium hexafluorophosphate (TBAPF₆), tetrabutylammonium perchlorate (TBAClO₄) and lithium perchlorate (LiClO₄). We note that the results presented here are not dependent on the identities of the solvent, supporting electrolyte or surface ligands used.

The nanocrystals were excited by a pulsed diode laser at a wavelength of 405 nm using low fluences (the average number of excitons per nanocrystal per pulse, $\langle N \rangle \leq 0.2$) to avoid multiexcitonic effects and to limit photocharging. The photoluminescence was collected confocally and sent to a Hanbury Brown/Twiss set-up (time resolution, 300 ps) to measure the second-order intensity correlation function, g_2 . The area of the central peak normalized to the area of a side peak is a measure of multiphoton emission probability during a single excitation cycle. Any $g_2(0)$ value less than 0.5 implies that the measured signal originates from a single quantum emitter (a single nanocrystal). For lifetime and blinking analyses, we used a time-tagged, time-resolved mode, in which we recorded the delay time of each photoluminescence photon with regard to the laser pulse. These data were analysed with the SYMPHOTIME software. All subsequent analysis and plotting were performed in ORIGIN 8.0.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 April; accepted 19 September 2011.

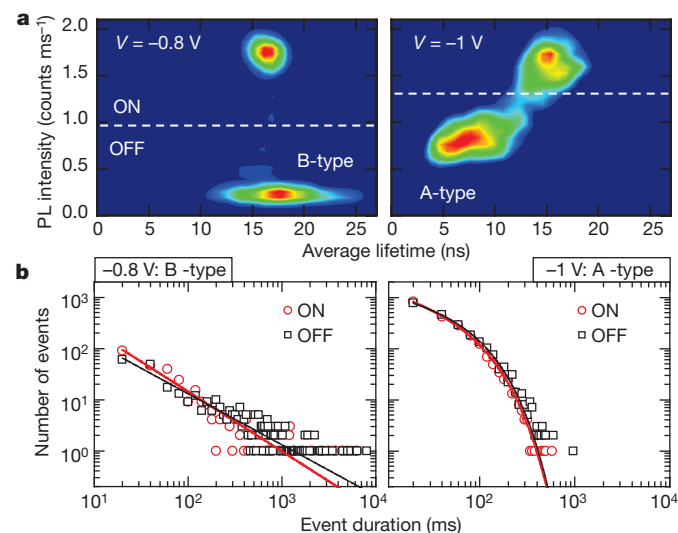


Figure 5 | Electrochemically controlled switching between distinct statistics for ON and OFF times in the same nanocrystal, accompanying the transition from B-type to A-type blinking. **a**, FLIDs indicating a nanocrystal switching from B-type blinking at -0.8 V (left) to A-type blinking at -1 V (right). Details of the analysis are given in Supplementary Fig. 9. **b**, Statistics for ON (red circles) and OFF (black squares) times for the FLIDs in **a**, in the log-log representation. At -0.8 V (B-type blinking), the data can be fitted to a power-law distribution, $\propto t^{-\alpha}$, with $\alpha = 1.17$ for the ON times (red line) and $\alpha = 1.00$ for the OFF times (black line). At -1 V (A-type blinking), this description is no longer valid; however, the data can be closely fitted by introducing an exponential cut-off such that the distribution is $\propto t^{-\alpha} \exp(-t/t_c)$, where $\alpha = 0.54$ and $t_c = 73.4$ ms for the ON times (red line) and $\alpha = 0.37$ and $t_c = 70.8$ ms for the OFF times (black line).

- Hoogenboom, J. P., Hernando, J., van Dijk, E. M. H. P., van Hulst, N. F. & García-Parajó, M. F. Power-law blinking in the fluorescence of single organic molecules. *ChemPhysChem* **8**, 823–833 (2007).
- Bout, D. A. V. et al. Discrete intensity jumps and intramolecular electronic energy transfer in the spectroscopy of single conjugated polymer molecules. *Science* **277**, 1074–1077 (1997).
- Riley, E. A., Bingham, C., Bott, E. D., Kahr, B. & Reid, P. J. Two mechanisms for fluorescence intermittency of single violamine R molecules. *Phys. Chem. Chem. Phys.* **13**, 1879–1887 (2011).
- Frantsuzov, P., Kuno, M., Janko, B. & Marcus, R. A. Universal emission intermittency in quantum dots, nanorods and nanowires. *Nature Phys.* **4**, 519–522 (2008).
- Nirmal, M. et al. Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature* **383**, 802–804 (1996).
- Fernando, D., Stefani, J. P. H. & Barkai, E. Beyond quantum jumps: blinking nanoscale light emitters. *Phys. Today* **6**, 34–39 (2009).
- Efros, A. L. & Rosen, M. Random telegraph signal in the photoluminescence intensity of a single quantum dot. *Phys. Rev. Lett.* **78**, 1110–1113 (1997).
- Klimov, V. I., Mikhailovsky, A. A., McBranch, D. W., Leatherdale, C. A. & Bawendi, M. G. Quantization of multiparticle Auger rates in semiconductor quantum dots. *Science* **287**, 1011–1013 (2000).
- Zhao, J., Nair, G., Fisher, B. R. & Bawendi, M. G. Challenge to the charging model of semiconductor nanocrystal fluorescence intermittency from off-state quantum yields and multiexciton blinking. *Phys. Rev. Lett.* **104**, 157403 (2010).
- Rosen, S., Schwartz, O. & Oron, D. Transient fluorescence of the off state in blinking CdSe/CdS/ZnS semiconductor nanocrystals is not governed by Auger recombination. *Phys. Rev. Lett.* **104**, 157404 (2010).
- Fisher, B. R., Eisler, H.-J., Stott, N. E. & Bawendi, M. G. Emission intensity dependence and single-exponential behavior in single colloidal quantum dot fluorescence lifetimes. *J. Phys. Chem. B* **108**, 143–148 (2004).
- Zhang, K., Chang, H., Fu, A., Alivisatos, A. P. & Yang, H. Continuous distribution of emission states from single CdSe/ZnS quantum dots. *Nano Lett.* **6**, 843–847 (2006).

13. García-Santamaría, F. *et al.* Breakdown of volume scaling in Auger recombination in CdSe/CdS heteronanocrystals: the role of the core–shell interface. *Nano Lett.* **11**, 687–693 (2011).
14. Gómez, D. E., van Embden, J., Mulvaney, P., Fernee, M. J. & Rubinsztein-Dunlop, H. Exciton–trion transitions in single CdSe–CdS core–shell nanocrystals. *ACS Nano* **3**, 2281–2287 (2009).
15. Jha, P. P. & Guyot-Sionnest, P. Trion decay in colloidal quantum dots. *ACS Nano* **3**, 1011–1015 (2009).
16. Houtepen, A. J. & Vanmaekelbergh, D. Orbital occupation in electron-charged CdSe quantum-dot solids. *J. Phys. Chem. B* **109**, 19634–19642 (2005).
17. Jha, P. P. & Guyot-Sionnest, P. Electrochemical switching of the photoluminescence of single quantum dots. *J. Phys. Chem. C* **114**, 21138–21141 (2010).
18. Chen, Y. *et al.* “Giant” multishell CdSe nanocrystal quantum dots with suppressed blinking. *J. Am. Chem. Soc.* **130**, 5026–5027 (2008).
19. Wang, X. *et al.* Non-blinking semiconductor nanocrystals. *Nature* **459**, 686–689 (2009).
20. Klimov, V. I., McGuire, J. A., Schaller, R. D. & Rupasov, V. I. Scaling of multiexciton lifetimes in semiconductor nanocrystals. *Phys. Rev. B* **77**, 195324 (2008).
21. McGuire, J. A. *et al.* Spectroscopic signatures of photocharging due to hot-carrier transfer in solutions of semiconductor nanocrystals under low-intensity ultraviolet excitation. *ACS Nano* **4**, 6087–6097 (2010).
22. Tisdale, W. A. *et al.* Hot-electron transfer from semiconductor nanocrystals. *Science* **328**, 1543–1547 (2010).
23. Li, S., Steigerwald, M. L. & Brus, L. E. Surface states in the photoionization of high-quality CdSe core/shell nanocrystals. *ACS Nano* **3**, 1267–1273 (2009).
24. Jeong, S. *et al.* Effect of the thiol–thiolate equilibrium on the photophysical properties of aqueous CdSe/ZnS nanocrystal quantum dots. *J. Am. Chem. Soc.* **127**, 10126–10127 (2005).
25. Hohng, S. & Ha, T. Near-complete suppression of quantum dot blinking in ambient conditions. *J. Am. Chem. Soc.* **126**, 1324–1325 (2004).
26. Fomenko, V. & Nesbitt, D. J. Solution control of radiative and nonradiative lifetimes: a novel contribution to quantum dot blinking suppression. *Nano Lett.* **8**, 287–293 (2008).
27. Jin, S., Song, N. & Lian, T. Suppressed blinking dynamics of single QDs on ITO. *ACS Nano* **4**, 1545–1552 (2010).
28. Tyagi, P. & Kambhampati, P. False multiple exciton recombination and multiple exciton generation signals in semiconductor quantum dots arise from surface charge trapping. *J. Chem. Phys.* **134**, 094706 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements C.G. and V.I.K. acknowledge support of the Center for Advanced Solar Photophysics, an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (BES). Y.G. and A.S. are supported by Los Alamos National Laboratory Directed Research and Development Fund. M.S., J.A.H. and H.H. are supported by NIH-NIGMS grant 1R01GM084702–01. This work was conducted, in part, at the Center for Integrated Nanotechnologies, a DOE/BES user facility.

Author Contributions C.G., M.S., J.A.H., V.I.K. and H.H. conceived the study. C.G., M.S. and H.H. designed the experiments. C.G. constructed the experimental set-up and performed the measurements under the guidance of M.S., V.I.K. and H.H. Y.G. synthesized and A.S. modified quantum dot materials under the guidance of J.A.H., C.G., V.I.K. and H.H. analysed and interpreted the data, and wrote the manuscript with the assistance of all other co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.I.K. (klimov@lanl.gov), H.H. (htoon@lanl.gov) or M.S. (sykoram@lanl.gov).

METHODS

Materials. Cadmium oxide, oleic acid (90%), 1-octadecene (ODE, 90%), 1-octadecane (OD, 90%), oleylamine, sulphur, selenium pellet and trioctylphosphine (TOP) were purchased from Aldrich and used without further purification. Trioctylphosphine oxide (TOPO, 90%) was purchased from Strem and used without further purification.

Nanocrystal synthesis. A 100-ml round-bottomed flask equipped with a reflux condenser and a thermocouple probe was charged with 1 g TOPO, 8 ml ODE and 0.38 mmol Cd-oleate under standard air-free conditions. The reaction system was evacuated for 30 min at room temperature (22 °C) and 30 min at 80 °C, and then the temperature was raised to 300 °C under argon. Following this, a mixture of 4 mmol of TOP-Se, 3 ml oleylamine and 1 ml of ODE was quickly injected into the reaction system. The temperature was then lowered to 270 °C for CdSe nanocrystal growth. After several minutes, the solution was cooled to room temperature and CdSe nanocrystals (diameter, $d = 3$ nm) were collected by precipitation with ethanol and centrifugation. The CdSe nanocrystals were redispersed in hexane.

The synthesis of core-shell CdSe/*n*CdS nanocrystals followed the successive ionic layer adsorption and reaction (SILAR) approach^{18,29,30} with modifications. A 250-ml round-bottomed flask was charged with $\sim 2 \times 10^{-7}$ mol pre-washed CdSe cores, 5 ml oleylamine and 5 ml OD. Here OD was chosen as the solvent because it alleviated the problem of precipitation observed during later stages of thick-shell growth. Elemental sulphur (0.2 M) dissolved in OD and 0.2 M Cd-oleate in ODE were used as precursors for shell growth. The quantity of precursor used for each addition of shell monolayer was calculated to account for the successive increases in particle volume as a function of increasing shell thickness. The reaction temperature was 240 °C and growth times were 1 h for sulphur and 2.5 h for Cd²⁺ precursors. Reactions were continued until a desired shell thickness was achieved. The core-shell nanocrystals were washed in a similar fashion as the CdSe cores, by precipitating two to three times with ethanol and redispersing in hexane. Relative photoluminescence quantum yields were determined by comparison with a standard dye (rhodamine 6G, 99%; Acros) and were observed to vary as a function of shell thickness. For the CdSe/9CdS nanocrystals used in the present study, the photoluminescence quantum yield was $\sim 30\%$. The purified core-shell nanocrystals were studied using transmission electron microscopy to determine their shapes and sizes and to confirm the growth of a thick, high-quality CdS shell over the CdSe core.

Ligand exchange. Core-shell nanocrystals were precipitated with ethanol then centrifuged for approximately 5 min (5,000 r.p.m., or 2,450g). The resulting pellet was redispersed in toluene. This procedure was repeated twice. The nanocrystal concentrations were calculated according to ref. 31. An amount of ligand (mercaptoundecanoic acid) equivalent to twice the number of moles of Cd-chalcogenide in the sample was added to the toluene solution. After 2 h, a solution of tetramethylammonium hydroxide in water (four times the number of moles of Cd-chalcogenide) was added dropwise. The nanocrystals were transferred from the toluene phase to the water phase. The water phase was separated from the toluene phase and precipitated with isopropanol, followed by centrifugation (~ 5 min at 5,000 r.p.m., or 2,450g). Finally, the pellet was redispersed in distilled water.

Electrochemical cell. A home-built electrochemical cell with a three-electrode configuration was used. As working electrode, we used an ITO-coated glass slide with sheet resistance of $\sim 50 \Omega$ (SPI Supplies). Before use, the electrode was sonicated in acetone and isopropanol baths, rinsed with deionized water, dried and plasma-etched for 10 min. The nanocrystals were directly deposited onto the electrode from a very dilute hexane or water solution. We note that plasma etching significantly improves the attachment of water-soluble nanocrystals by providing a hydrophilic surface. As a counterelectrode, we used platinum gauze attached to a platinum wire. The high-surface-area gauze was used to achieve uniform current density across the working electrode. A silver wire was used as a quasi-reference electrode. This electrode was calibrated using the Ru^{3+/2+} redox-couple of [Ru(bpy)₃](PF₆)₂ (refs 32, 33). By comparison of the half-wave potentials obtained with the silver wire, standard calomel electrode and Ag/Ag(NO₃) reference electrodes, we found that the silver quasi-reference is offset from the normal

hydrogen electrode by 0.31 ± 0.01 V. All potentials reported in the main text are measured relative to the silver quasi-reference. The electrochemical experiments were performed using several combinations of solvents (acetonitrile and propylene carbonate) and supporting electrolytes (all concentrations, 0.1 M): tetrabutylammonium hexafluorophosphate (TBAPF₆), tetrabutylammonium perchlorate (TBAClO₄) and lithium perchlorate (LiClO₄). The results presented here are not dependent on the identities of the solvent, supporting electrolyte or surface ligands used.

Optical set-up. The excitation source was a PicoQuant pulsed diode laser producing ~ 30 -ps pulses at 405 nm with a repetition rate of 2.5–40 MHz. Most of the experiments were performed at 2.5 MHz, which corresponds to pulse-to-pulse separation of 400 ns, an order of magnitude greater than the longest photoluminescence lifetimes. This allows us to minimize ‘pile-up’ effects and parasitic charge accumulation due to possible photocharging. The average nanocrystal excitonic occupancies generated per pulse, $\langle N \rangle$, were estimated from absorption cross-sections calculated using nanocrystal sizes derived from transmission electron microscopy data and were independently verified by photoluminescence saturation and intensity-dependent g_2 measurements³⁴. Photoluminescence was excited and collected through an oil-immersion Olympus objective with a numerical aperture of 1.3. After reflection from a dichroic mirror (Semrock), photoluminescence then went through a long-pass or band-pass filter (Semrock). A flip mirror was used to send emission to a 500-mm spectrometer equipped with a liquid-nitrogen-cooled silicon charge-coupled device. Emission from the nanocrystals typically peaked around 620 nm with a full-width of ~ 25 nm at half maximum. A Hanbury Brown/Twiss set-up was realized using a 50/50 beam splitter and two avalanche photodiodes (APDs; SPCM-AQRH-14, Perkin Elmer) with a quantum efficiency of $\sim 50\%$ at the photoluminescence wavelength, a time jitter of ~ 300 ps and a dark count rate of < 100 Hz. The single-photon counting device was a PicoHarp 300 stand-alone module (PicoQuant). Two APDs were used to produce start and stop signals in the measurements of the second-order intensity correlation function, whereas the synchronization pulse of the laser provided the start signal in the time-tagged, time-resolved mode. Photon arrival times were recorded from one of the APDs (stop signal).

Analysis. For the analysis of raw time-tagged, time-resolved data, we used the SYMPHOTIME software. All subsequent analysis and plotting were performed in ORIGIN 8.0. For the dynamical correlated lifetime-intensity analysis, we chose a bin time, corresponding to more than 100 photons per bin on average, to ensure a reliable bi-exponential fitting for each decay curve. We fixed the lifetimes on the basis of the values produced by the global fit procedure and constrained the amplitudes to be positive numbers. To enhance the precision, we used a Poissonian maximum-likelihood estimator. To confirm the validity of the multi-exponential approach, we also constructed FLIDs for which the lifetime for each bin was calculated as a weighted average of photoluminescence photon arrival times, that is, without any fitting procedure. The resulting FLIDs were similar to those produced by a multi-exponential fit, as illustrated in Supplementary Fig. 13.

29. Li, J. J. *et al.* Large-scale synthesis of nearly monodisperse CdSe/CdS core/shell nanocrystals using air-stable reagents via successive ion layer adsorption and reaction. *J. Am. Chem. Soc.* **125**, 12567–12575 (2003).
30. Vela, J. *et al.* Effect of shell thickness and composition on blinking suppression and blinking mechanism in ‘giant’ CdSe/CdS nanocrystal quantum dots. *J. Biophoton.* **3**, 706–717 (2010).
31. Yu, W. W., Qu, L., Guo, W. & Peng, X. Experimental determination of the extinction coefficient of CdTe, CdSe, and CdS nanocrystals. *Chem. Mater.* **15**, 2854–2860 (2003).
32. Juris, A. *et al.* Ru(II) polypyridine complexes: photophysics, photochemistry, electrochemistry, and chemiluminescence. *Coord. Chem. Rev.* **84**, 85–277 (1988).
33. DeLaive, P. J., Foreman, T. K., Giannotti, C. & Whitten, D. G. Photoinduced electron transfer reactions of transition-metal complexes with amines. Mechanistic studies of alternate pathways to back electron transfer. *J. Am. Chem. Soc.* **102**, 5627–5631 (1980).
34. Park, Y. S. *et al.* Near-unity quantum yields of biexciton emission from CdSe/CdS nanocrystals measured using single-particle spectroscopy. *Phys. Rev. Lett.* **106**, 187401 (2011).

Electrically driven directional motion of a four-wheeled molecule on a metal surface

Tibor Kudernac^{1,2,†*}, Nopporn Ruangsapichat^{2*}, Manfred Parschau³, Beatriz Maciá¹, Nathalie Katsonis^{1,2,†}, Syuzanna R. Harutyunyan¹, Karl-Heinz Ernst^{3,4} & Ben L. Feringa^{1,2}

Propelling single molecules in a controlled manner along an unmodified surface remains extremely challenging because it requires molecules that can use light, chemical or electrical energy to modulate their interaction with the surface in a way that generates motion. Nature's motor proteins^{1,2} have mastered the art of converting conformational changes into directed motion, and have inspired the design of artificial systems³ such as DNA walkers^{4,5} and light- and redox-driven molecular motors^{6–11}. But although controlled movement of single molecules along a surface has been reported^{12–16}, the molecules in these examples act as passive elements that either diffuse along a preferential direction with equal probability for forward and backward movement or are dragged by an STM tip. Here we present a molecule with four functional units—our previously reported rotary motors^{6,8,17}—that undergo continuous and defined conformational changes upon sequential electronic and vibrational excitation. Scanning tunnelling microscopy confirms that activation of the conformational changes of the rotors through inelastic electron tunnelling propels the molecule unidirectionally across a Cu(111) surface. The system can be adapted to follow either linear or random surface trajectories or to remain stationary, by tuning the chirality of the individual motor units. Our design provides a starting point for the exploration of more sophisticated molecular mechanical systems with directionally controlled motion.

Figure 1a shows the molecular system and the distinctive design features that allow it to move upon electronic excitation in a preferred and linear direction across a surface (see Supplementary Information for synthesis and characterization of the molecules). The molecule has four chiral units based on the unidirectional rotary motors (Fig. 1b) that have been studied in solution^{6,8} and liquid crystalline phases¹⁷. As illustrated in Fig. 1c, the rotary motors undergo geometric changes as a result of sequential configurational and conformational switching induced by electronic and vibrational excitation, respectively.

After sublimation onto a Cu(111) surface, individual molecules were imaged with a scanning tunnelling microscope (STM) at 7 K (see Supplementary Fig. 1 for details of the measuring and manipulation procedure). Imaging conditions are sufficiently mild that no changes are induced upon continuous scanning (Supplementary Fig. 2), with single molecules appearing as three or four joined-together lobes (each lobe corresponds to a rotary motor unit). Excitation can vary the STM contrast of individual stereoisomers between three and four lobe structures, reflecting excitation-induced geometrical changes in the molecule that lead to different conformers; but the original molecular feature always reappears (see Supplementary Information and Supplementary Figs 3 and 4 for details).

If the four motors all rotate in the same direction, linear translational motion of the molecule across the surface should occur through a paddlewheel type movement (Fig. 1d and e). Of all the possible isomers, this requirement is only met by the *meso*-(*R,S,R,S*)

isomer sketched in Fig. 1a. Figure 2a shows a STM image of one such molecule on a Cu(111) surface, at its starting position. The STM tip is then positioned above the centre of the molecule to apply a voltage pulse larger than 500 mV, to induce movement that is observed by re-scanning the area under mild STM imaging conditions. Ten excitation steps result in the molecule travelling 6 nm across the surface along the trajectory indicated in Fig. 2b to the final position shown in Fig. 2c (see also Supplementary Movie 1). Figure 2e shows a selection of STM images recorded between electrically induced movements. (For additional examples, see Supplementary Fig. 6a.) We note that although backward movement is never observed, the trajectory is not perfectly straight because achieving simultaneous excitation of all motor units during each step is not trivial. Nonetheless, the data in Fig. 2 clearly show that the action of the motor units within the *meso*-(*R,S,R,S*) isomer induces directional propulsion of the molecule along the surface (see also Supplementary Information for a statistical analysis of the linearity of the translation and Supplementary Fig. 5).

In solution, unidirectional rotation of the motor units involves helix inversions and double bond isomerizations (Fig. 1c). On Cu(111), we anticipate that the propulsion of the molecule should also be the direct consequence of a combination of these two processes. In this context, we note that the conformation of porphyrins^{18,19} and biphenyl molecules adsorbed on surfaces²⁰ has been switched through STM-induced vibrational excitation. Similarly, configurational switching (isomerization) of azobenzene²¹ and 2-butene²² on surfaces has been realized by exciting electronic transitions with tunnelling electrons in a manner analogous to excitation by a photon²³.

Our view of the configurational and conformational changes resulting in the movement of a single *meso*-(*R,S,R,S*) molecule along the surface is sketched in Fig. 1e. Electronic excitation induces double-bond isomerization that is followed by helix inversion (see Fig. 1c), resulting in paddlewheel-like motion pushing the entire molecule one step forward. Movement is of course possible only if the adsorption energy of the molecule is less than the energy released in the isomerization step upon excitation. Because the motor units are highly strained helical systems, their aromatic parts cannot align parallel to the surface (Fig. 1e) but are oriented edge-on and physisorbed through their hydrogen atoms. This leads to binding energies for individual motor units considerably below the 0.5 eV observed with, for example²⁴, parallel benzene on Cu(111). Finally, we note that one paddlewheel-like motion corresponds to 0.7 nm of forward movement and that ten excitation events should thus result in the molecule having traversed approximately 7 nm. In practice, we observe near-ideal linear movement resulting in 6 nm after ten steps (Fig. 2b).

In addition to changing their positions, molecules also exhibit a modified STM contrast after movement. Irreversible changes in the structure of the molecule, however, are excluded, as the contrast changes are fully reversible. To explore this effect further, the STM

¹Centre for Systems Chemistry, Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. ²Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. ³Nanoscale Materials Science, Empa, Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, CH-8600 Dübendorf, Switzerland. ⁴Department of Chemistry, University Zurich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [†]Present address: MESA+ Institute for Nanotechnology, University of Twente, PO Box 217, 7500 AE, Enschede, The Netherlands.

*These authors contributed equally to this work.

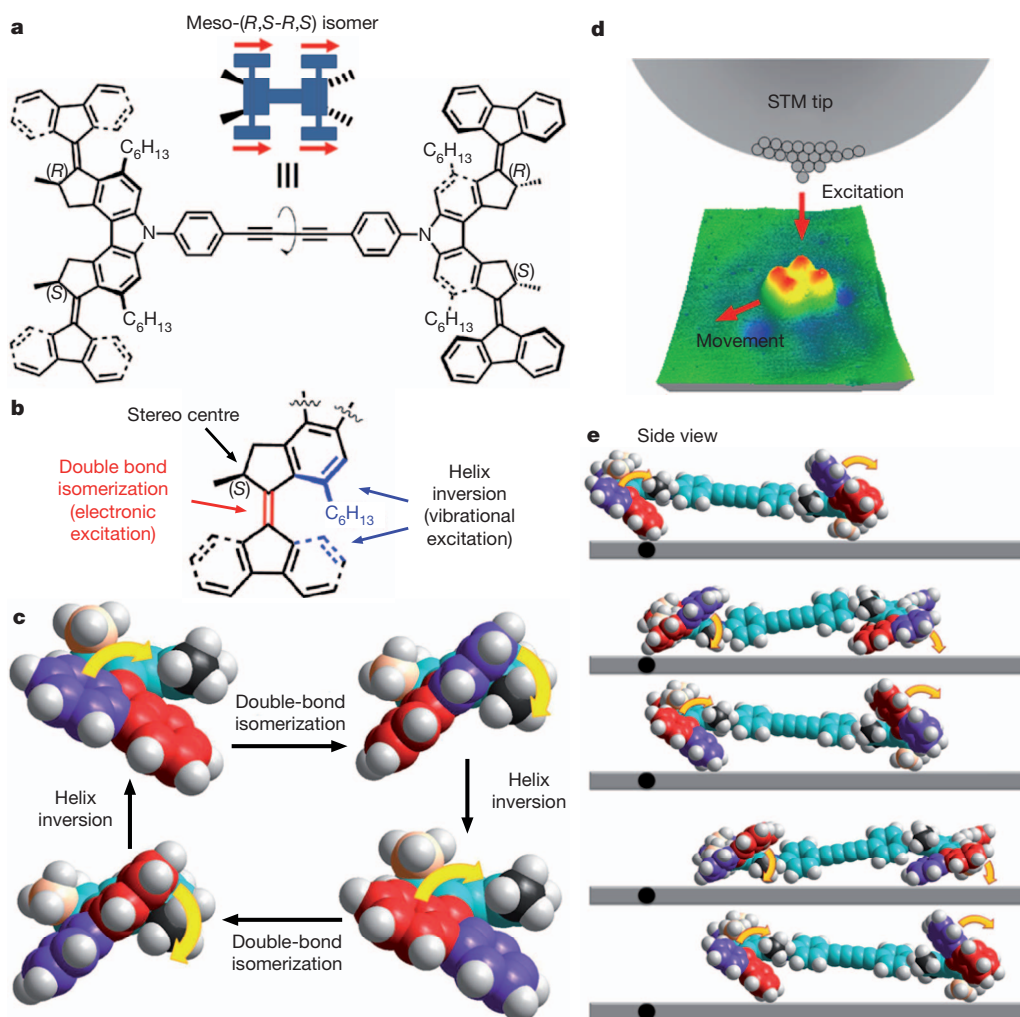


Figure 1 | Structure of the four-wheeled molecule. **a**, Structure and cartoon representation of the *meso*-(*R,S-R,S*) isomer. Red arrows indicate the direction in which the rotary action of the individual motor units propels the molecule. (*R*) and (*S*) indicate the absolute configurations at the stereogenic centres. The black solid and dashed wedges of the cartoon indicate the orientations of the methyl groups, respectively. **b**, Structural details of the rotary motor unit. The double bond (red) functions as the axle in rotation and undergoes *trans*-to-*cis* isomerization when electronically excited. Interconversion between helical conformers, arising from steric overcrowding in the region highlighted in blue, is achieved by vibrational excitation. The stereocentre in the cyclopentane ring

determines the stability of each conformer and the direction of rotation of the motor. **c**, Schematic representation of the 360° rotation of the rotary motor, involving two double-bond isomerization steps and two helix inversion steps. Different colours in the model highlight the different positions of atoms during the rotor action. For clarity, the hexyl groups are substituted by methyl groups. **d**, Schematic representation of the experiment. The bias voltage U is applied to the sample. Electrons tunnelling through the molecule excite vibrational and electronic states and induce translational movement on the surface. **e**, Molecular model representation (side view) of the paddlewheel-like motion of the four-wheeled molecule (see also Supplementary Movie 4).

tip was positioned above one lobe of the molecule and voltage pulses of between 200 and 350 mV were applied to induce contrast modification without displacement (Fig. 3a and b; Supplementary Movie 2). The threshold for inducing conformational changes within the motor units, without movement, is 200 mV. This is illustrated by the action spectrum in Fig. 3c, which is related to the vibrational density of states²⁵, with the threshold bias of 200 mV corresponding to the excitation of the C=C double bond stretching vibration²⁶ (204 meV, $\sim 1,600\text{ cm}^{-1}$) that induces the helix inversion of the rotor unit.

Another insight into the mechanism of propulsion is provided by the polarity dependence of our experiments. Figure 3d shows that negative pulses of $-1,500\text{ mV}$ induce contrast modifications but no movement. We note that vibrational excitation by inelastically tunnelling electrons is polarity independent. In contrast, a positive pulse of $+1,500\text{ mV}$, applied with the same tip on the same molecule and in the same position, induces a reproducible translation of the molecule (Fig. 3d). The polarity dependence of the movement points clearly to the electronic origin of the motor rotation. The corresponding action spectrum (Fig. 2d) of the molecule suggests the presence of the

lowest unoccupied molecular orbital (LUMO) at around 500 meV above the Fermi level. Resonant tunnelling into the LUMO of the molecule will lead to transient excitation of the molecule by forming a negative ion resonance²⁷. This results in excitation analogous to light-induced double-bond isomerization.

We excluded from our experiments all cases of dragging^{14–16} or pick-up-and-drop²⁸ with the STM tip, and random or directional thermal diffusion^{12,13} of molecules was not observed under the experimental conditions in any of the measurements. We also exclude vibrationally induced lateral hopping (activated diffusion) as a cause for the movement because the minimum energy of 500 mV needed to induce it is well above relevant vibrational energies²⁹. Instead, the polarity dependence and the very low tunnelling currents during manipulation support an electronic excitation mechanism.

Additional experimental corroboration for electrically driven directional motion comes from two series of experiments probing the impact of the chirality and geometry of the molecule on its motion (Fig. 4). A key point is that improper orientation ('wrong landing') of the *meso*-isomer on the surface completely changes its behaviour

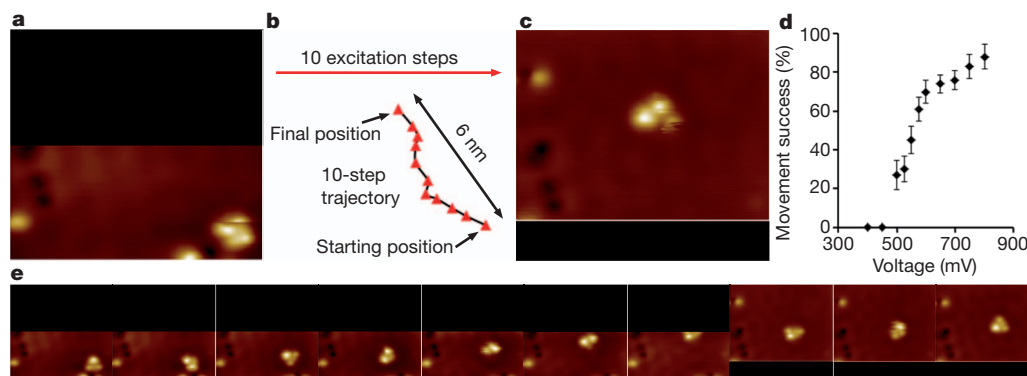


Figure 2 | Linear movement of the *meso*-(*R,S*-*R,S*) isomer. **a**, STM image (imaging parameters: area $10.2 \text{ nm} \times 9.3 \text{ nm}$, current $I = 74 \text{ pA}$, $U = 47 \text{ mV}$) of the initial position. The black area was scanned only after the molecule moved into it. **b**, Trajectory depicting the individual steps taken (see Supplementary Movie 1). **c**, Final position after ten consecutive voltage pulses. **d**, The action spectrum for

movement shows a voltage threshold at 500 mV. Each data point represents 8 to 40 manipulations performed on various molecules ($I = 30\text{--}50 \text{ pA}$). Error bars represent the standard deviation from the probability for successful events (see equation (1) in the Supplementary Information). **e**, STM frames corresponding to individual steps of the trajectory in **b** excluding starting and final position.

(Fig. 4a–c): the changes in contrast arising from helix inversions occur (Supplementary Fig. 6b), but the molecule shows no translational movement upon manipulation. In the second series of experiments, individual (*R,R-R,R*) or (*S,S-S,S*) enantiomers of the racemate adsorbed on the surface were found to spin and randomly move across the surface as illustrated in Fig. 4d–f, which shows a molecule that travels an overall distance of only 3 nm after 40 excitation steps (see also Supplementary Figs 5 and 6c for STM image frames and

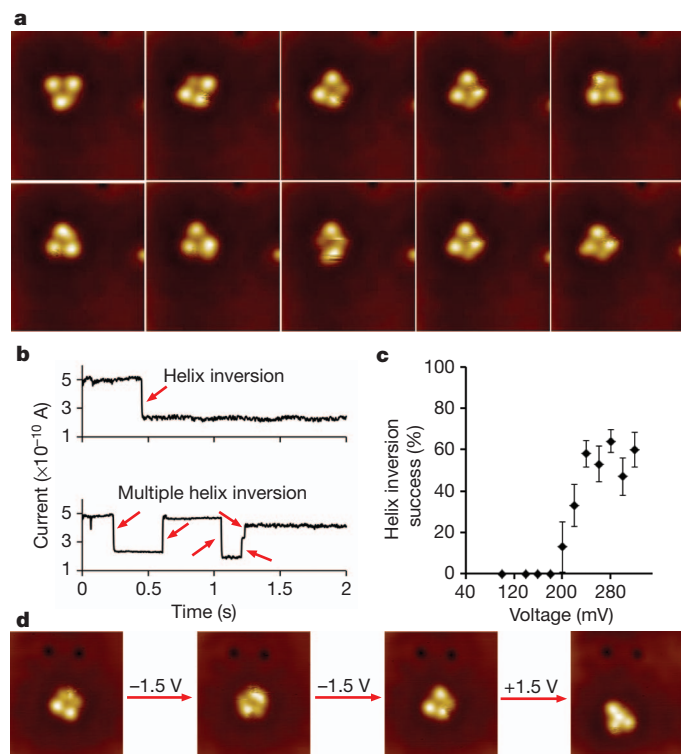


Figure 3 | Helix inversion at lower bias voltage and polarity dependence of propulsion. **a**, Voltage pulses between 200 and 350 mV lead to conformational changes but no movement ($7.0 \text{ nm} \times 7.8 \text{ nm}$, $I = 43 \text{ pA}$, $U = 47 \text{ mV}$). **b**, Steps in the real-time traces (red arrows) of the tunnelling current ($U = 200 \text{ mV}$) indicate the internal conformational changes of the molecule. **c**, The threshold voltage for helix inversion is 200 mV (>200 manipulations, $I = 30\text{--}50 \text{ pA}$). Error bars represent the standard deviation from the probability for successful events (see equation (1) of the Supplementary Information). **d**, Negative bias (tunnelling from the sample to the tip) induces contrast changes but no movement ($7.0 \text{ nm} \times 7.8 \text{ nm}$, $I = 43 \text{ pA}$, $U = 47 \text{ mV}$). Positive voltage pulses lead to movement and helix inversion.

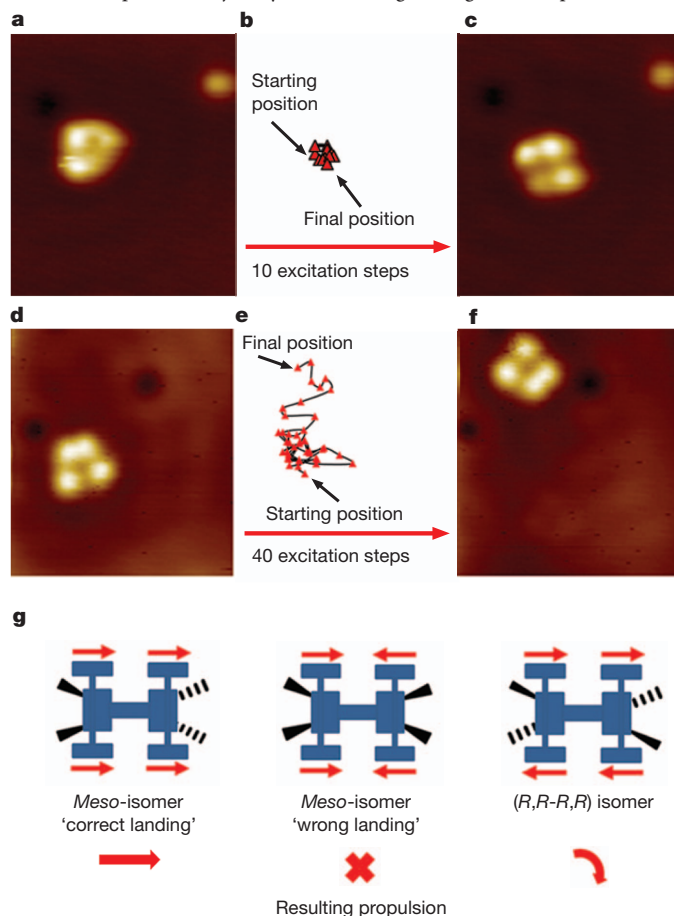


Figure 4 | Control over motion by the geometries of the four motors.

a–c, Absence of motion of a ‘wrongly landed’ *meso*-isomer. **a**, STM image of a ‘wrongly landed’ *meso*-isomer ($7.0 \text{ nm} \times 7.8 \text{ nm}$, $I = 74 \text{ pA}$, $U = 47 \text{ mV}$). **b**, Plot of all positions occupied by this *meso*-isomer after each manipulation (see Supplementary Fig. 6b for individual STM images). **c**, STM image of the molecule after ten voltage pulses. **d–f**, Random motion of a single (*R,R-R,R*) or (*S,S-S,S*) enantiomer. **d**, STM image at initial position ($7.0 \text{ nm} \times 7.8 \text{ nm}$, $I = 47 \text{ pA}$, $U = 47 \text{ mV}$). **e**, Trajectory of 40 individual steps (see Supplementary Movie 3 and Supplementary Fig. 6c for individual STM images). **f**, STM image at final position. **g**, Sketch of directionality of movement induced by concerted rotation of the motor units. Two distinct ‘landing geometries’ of the *meso*-isomer lead to either directional movement or to no movement at all. (*R,R-R,R*) or (*S,S-S,S*) isomers move randomly.

Supplementary Movie 3). We note also that in contrast to the ‘correctly landed’ *meso*-isomer (Fig. 2), the STM appearance is dominated by four lobes (Fig. 4; for details of the STM appearance see Supplementary Fig. 4).

As sketched in Fig. 4g, these differences in motion are readily explained by the chirality and geometry of the molecule on the surface. In the case of the *meso*-isomer, two geometries need to be considered because free rotation around the bisalkyne C–C single bond of the frame is locked upon physisorption (Figs 1a and 4g). When the *meso*-(*R,S,R,S*) isomer is adsorbed on the surface in the proper orientation (‘correct landing’), the four motor units can act in concert so that conrotatory motion moves the molecule along. In contrast, in the ‘wrongly landed’ *meso*-(*R,S-R,S*) isomer the combined effects of the motor units cancel out and thus preclude translational movement (Fig. 4a–c and g). And in the (*R,R-R,R*) or (*S,S-S,S*)-enantiomers of the racemic mixture, the motor units on opposite sides of the molecule will rotate in a disrotatory fashion and, in an ideal case, cause these molecules to spin; in a non-ideal case, the molecules exhibit random translational motion in addition to the spinning motion (Fig. 4d–f and g).

The directional and exclusively forward movement of the ‘correctly landed’ *meso*-isomer, the lack of movement of the ‘wrongly landed’ *meso*-isomer and the random movement of individual molecules of the racemic isomer are direct consequences of the molecular design and provide compelling evidence that the translational movement originates from the concerted action of the molecular motor units (Fig. 4g). A single molecule with intrinsic motor functions is thus capable of converting an external energy input into unidirectional movement along a surface.

METHODS SUMMARY

STM manipulations were performed with a custom-built low-temperature STM under ultrahigh vacuum conditions with base pressure below 10^{-10} mbar at a temperature of 7 K. To ensure that no molecule manipulation occurs upon scanning, constant-current STM images were recorded with positive voltages below 60 mV and tunnelling currents between 30 and 100 pA. The Cu(111) surface was cleaned by 30 min of argon sputtering followed by 15 min of annealing at 820 K. Molecules were sublimed at 553 K for 2 min on the pre-cooled clean surface held at 40 K. For single-molecule manipulation (see Supplementary Information) the STM tip was positioned over the molecule and the feedback loop was then switched off. A voltage pulse between the steady STM tip and the surface was applied for a few seconds while recording the tunnelling current. A step-like variation in tunnelling current indicated a manipulation step and rescanning of the area revealed the type of manipulation. Determining the threshold voltages for the helix inversion and propulsion was performed with currents between 30 and 50 pA.

Received 10 January; accepted 15 September 2011.

- Schliwa, M. & Woehlke, G. Molecular motors. *Nature* **422**, 759–765 (2003).
- van den Heuvel, M. G. L. & Dekker, C. Motor proteins at work for nanotechnology. *Science* **317**, 333–336 (2007).
- Ismagilov, R. F., Schwartz, A., Bowden, N. & Whitesides, G. M. Autonomous movement and self-assembly. *Angew. Chem. Int. Edn Engl.* **41**, 652–654 (2002).
- Lund, K. *et al.* Molecular robots guided by prescriptive landscapes. *Nature* **465**, 206–210 (2010).
- Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202–205 (2010).
- Browne, W. R. & Feringa, B. L. Making molecular machines work. *Nature Nanotechnol.* **1**, 25–35 (2006).

- Kay, E. R., Leigh, D. A. & Zerbetto, F. Synthetic molecular motors and mechanical machines. *Angew. Chem. Int. Edn Engl.* **46**, 72–191 (2007).
- Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. Light-driven molecular rotor. *Nature* **401**, 152–155 (1999).
- Balzani, V., Credi, A., Raymo, F. M. & Stoddart, J. F. Artificial molecular machines. *Angew. Chem. Int. Edn Engl.* **39**, 3348–3391 (2000).
- van Delden, R. A. *et al.* Unidirectional molecular motor on a gold surface. *Nature* **437**, 1337–1340 (2005).
- Green, J. E. *et al.* A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimeter. *Nature* **445**, 414–417 (2007).
- Shirai, Y., Osgood, A. J., Zhao, Y., Kelly, K. F. & Tour, J. M. Directional control in thermally driven single-molecule nanocars. *Nano Lett.* **5**, 2330–2334 (2005).
- Wong, K. L. *et al.* A molecule carrier. *Science* **315**, 1391–1393 (2007).
- Gimzewski, J. K. & Joachim, C. Nanoscale science of single molecules using local probes. *Science* **283**, 1683–1688 (1999).
- Grill, L. *et al.* Rolling a single molecular wheel at the atomic scale. *Nature Nanotechnol.* **2**, 95–98 (2007).
- Chiaravalloti, F. *et al.* A rack-and-pinion device at the molecular scale. *Nature Mater.* **6**, 30–33 (2007).
- Eelkema, R. *et al.* Nanomotor rotates microscale objects. *Nature* **440**, 163 (2006).
- Iancu, V., Deshpande, A. & Hla, S.-W. Manipulating Kondo temperature via single molecule switching. *Nano Lett.* **6**, 820–823 (2006).
- Qiu, X. H., Nazin, G. V. & Ho, W. Mechanisms of reversible conformational transitions in a single molecule. *Phys. Rev. Lett.* **93**, 196806 (2004).
- Lastapis, M. *et al.* Picometer-scale electronic control of molecular dynamics inside a single molecule. *Science* **308**, 1000–1003 (2005).
- Aleman, M. *et al.* Adsorption and switching properties of azobenzene derivatives on different noble metal surfaces: Au(111), Cu(111), and Au(100). *J. Phys. Chem. C* **112**, 10509–10514 (2008).
- Sainoo, Y., Kim, Y., Komeda, T., Kawai, M. & Shigekawa, H. Observation of *cis*-2-butene molecule on Pd(110) by cryogenic STM: site determination using tunnelling-current-induced rotation. *Surf. Sci.* **536**, L403–L407 (2003).
- Henzl, J. & Morgenstern, K. An electron induced two-dimensional switch made of azobenzene derivatives anchored in supramolecular assemblies. *Phys. Chem. Chem. Phys.* **12**, 6035–6044 (2010).
- Lukas, S., Vollmer, S., Witte, G. & Wöll, C. Adsorption of acenes on flat and vicinal Cu(111) surfaces: step induced formation of lateral order. *J. Chem. Phys.* **114**, 10123–10130 (2001).
- Ueba, H. & Persson, B. N. Action spectroscopy for single-molecule motion induced by vibrational excitation with a scanning tunneling microscope. *Phys. Rev. B* **75**, 041403 (2007).
- Parschau, M., Passerone, D., Rieder, K.-H., Hug, H. J. & Ernst, K.-H. Switching the chirality of a single adsorbate. *Angew. Chem. Int. Edn Engl.* **48**, 4065–4068 (2009).
- Seideman, T. Current-driven dynamics in molecular-scale devices. *J. Phys. Condens. Matter* **15**, R521, doi:10.1088/0953-8984/15/14/201 (2003).
- Eigler, D. M., Lutz, C. P. & Rudge, W. E. An atomic switch realized with the scanning tunnelling microscope. *Nature* **352**, 600–603 (1991).
- Parschau, M., Rieder, K.-H., Hug, H. J. & Ernst, K.-H. Single-molecule chemistry and analysis: mode-specific dehydrogenation of adsorbed propene by inelastic electron tunneling. *J. Am. Chem. Soc.* **133**, 5689–5691 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the Netherlands Organization for Scientific Research (NWO-CW) (B.L.F. and T.K. through a VENI grant), the Swiss Secretary for Education and Research and the Swiss National Science Foundation (K.-H.E. and M.P.), and the European Research Council (ERC advanced grant 227897 to B.L.F.).

Author Contributions N.R., B.M., S.R.H. and B.L.F. designed the four-wheeled molecule and N.R. conducted its synthesis and characterization. T.K., M.P., N.K. and K.-H.E. designed the STM experiments and contributed to their interpretation. T.K. and M.P. performed the STM experiments at Empa. T.K., K.-H.E. and B.L.F. wrote the manuscript. S.R.H., K.-H.E. and B.L.F. conceived and guided the research. All authors discussed the results and implications and commented on the manuscript at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.R.H. (s.harutyunyan@rug.nl), K.H.E. (karl-heinz.ernst@empa.ch) or B.L.F. (b.l.feringa@rug.nl).

A long-lived lunar dynamo driven by continuous mechanical stirring

C. A. Dwyer¹, D. J. Stevenson² & F. Nimmo¹

Lunar rocks contain a record of an ancient magnetic field that seems to have persisted for more than 400 million years^{1,2} and which has been attributed to a lunar dynamo^{3,4}. Models of conventional dynamos driven by thermal or compositional convection have had difficulty reproducing the existence and apparently long duration of the lunar dynamo^{5–7}. Here we investigate an alternative mechanism of dynamo generation: continuous mechanical stirring arising from the differential motion, due to Earth-driven precession of the lunar spin axis, between the solid silicate mantle and the liquid core beneath^{8,9}. We show that the fluid motions and the power required to drive a dynamo operating continuously for more than one billion years and generating a magnetic field that had an intensity of more than one microtesla 4.2 billion years ago³ are readily obtained by mechanical stirring. The magnetic field is predicted to decrease with time and to shut off naturally when the Moon recedes far enough from Earth that the dissipated power is insufficient to drive a dynamo; in our nominal model, this occurred at about 48 Earth radii (2.7 billion years ago). Thus, lunar palaeomagnetic measurements may be able to constrain the poorly known early orbital evolution of the Moon. This mechanism may also be applicable to dynamos in other bodies, such as large asteroids.

Several lines of evidence^{2,10} point to the Moon having an iron core that is 300–400 km in radius and probably at least partly molten at the present day. A recent palaeomagnetic study³ has strengthened the case that the observed lunar magnetic anomalies are due to an ancient dynamo⁴ rather than an external source such as impacts¹¹. Although deriving palaeointensities from lunar samples is difficult^{2,12}, the results of ref. 3 raise the question of how a lunar dynamo could be maintained.

Dynamos are usually assumed to arise as a result of thermal or compositional convection driven by removal of heat into the overlying mantle¹³. A long-lived, convection-driven lunar dynamo is difficult to sustain because it requires rapid core cooling. The lunar mantle is unlikely to permit persistent rapid cooling, unless special conditions are invoked^{5–7}.

Shortly after the Moon's formation, when the Earth–Moon distance was <26–29 Earth radii (R_E), the lunar core precessed with the mantle^{14,15}, but as the distance increased, differential motion between core and mantle is predicted to have occurred because the core and the mantle precess around slightly different axes^{8,9,14,15} (Fig. 1, inset). The essential idea in our model is that this leads to turbulent motions in the core with amplitudes of the order of the core–mantle differential velocity. These motions are of sufficient amplitude to produce a magnetic Reynolds number of $\sim 10^4$ or more, which is a necessary but not sufficient condition for a dynamo (Supplementary Information, section 2). Our analysis is consistent with ref. 16, where it was shown that differential motion can drive a dynamo under some conditions; and in Supplementary Information, section 2, we argue that the early lunar core satisfies these conditions.

The turbulent friction from the differential motion at the core–mantle boundary (CMB), results in a small departure of the Moon's rotation axis from the plane defined by the lines perpendicular to the ecliptic and the plane of the lunar orbit^{8,9}. The power dissipated at this

boundary at the present day is $(6.0 \pm 1.6) \times 10^7$ W (ref. 9). In the weak (low-viscosity) limit, in which core–mantle coupling is small, the power varies as $n^3 \sin^3(I_e)$ (ref. 9), where n is the mean motion of the lunar orbit and I_e (the equatorial inclination) is the angle between the lunar spin axis and the ecliptic normal (Fig. 1, inset).

The power available to produce a magnetic field depends on how both n (or equivalently, the semimajor axis of the lunar orbit, a) and I_e have varied over time (Methods). Because the Moon is in a Cassini state¹⁷, the lunar semimajor axis and equatorial inclination are related¹⁸. At early times, when the Moon was closer to Earth, the equatorial inclination was larger (Supplementary Fig. 1). The Moon

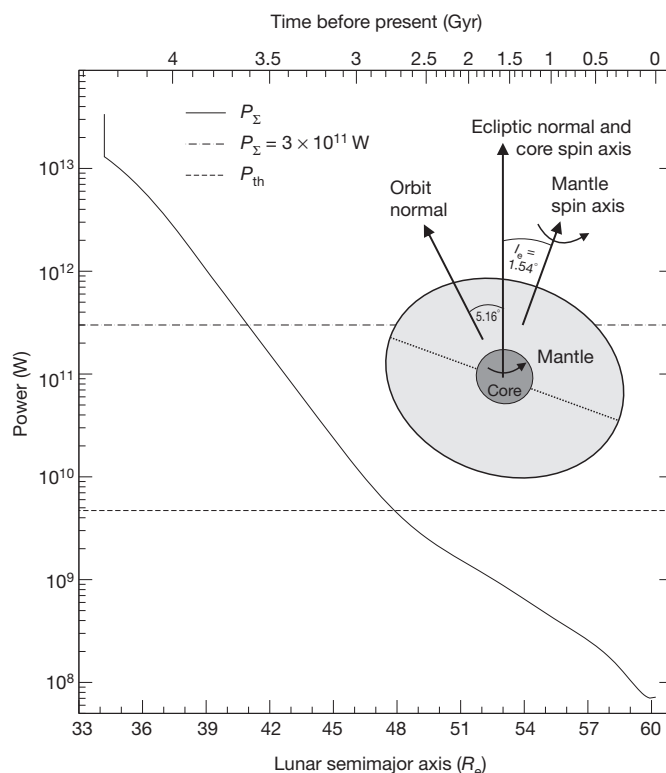


Figure 1 | The total power deposited into the lunar core. The solid line shows the total power deposited into the lunar core, P_Σ , as a function of semimajor axis (equation (1); lower x axis). The adiabatic threshold value is marked P_{th} (dashed line). The dash-dot line is at $P_\Sigma = 3 \times 10^{11}$ W (see text). The time before present is plotted at the top, calculated using the results of ref. 20 (Methods). There are 7×10^{28} J of energy available for dynamo generation between the time of the Cassini-state transition and the cessation of the dynamo (using our nominal evolution model). The inset shows the geometry of the situation. The mean motion, n , of an orbit is related to the semimajor axis, a , through $n = (GM_E/a^3)^{0.5}$, where M_E is the mass of Earth. For the purposes of display, the calculations were performed with no truncation of coefficients.

¹Department of Earth and Planetary Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. ²Division of Geologic and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

probably also underwent a transition in the Cassini state it was in at $a \approx 34R_e$ (ref. 18).

The temporal evolution of the lunar orbit is poorly constrained at times before 600 million years (0.6 Gyr) ago¹⁹. Here we have used the results of ref. 20 to relate the lunar semimajor axis to the time before present (Methods and Supplementary Figs 1 and 2); we discuss details and how the use of other models would affect our results in Supplementary Information, section 3. On the basis of the measured present-day dissipation rate⁹, the total power dissipated at the lunar CMB, P_Σ , at time t is given by

$$P_\Sigma(t) = \left(\frac{a(t_n)}{a(t)}\right)^{9/2} \left(\frac{\sin[I_e(t)]}{\sin[I_e(t_n)]}\right)^3 P_\Sigma(t_n) \quad (1)$$

$$\approx 3 \times 10^{20} \text{ W} \times \frac{\sin^3[I_e(t)]}{(a(t)/R_e)^{9/2}}$$

where t_n denotes the time at the present day and $I_e(t_n) = 1.54^\circ$.

Not all of the power dissipated at the CMB is available to generate a dynamo; thus, there is a threshold power, P_{th} , at which the dynamo will cease. The power available to generate a dynamo is

$$P_{dyn} = P_\Sigma - P_{th} \quad (2)$$

However, the value of P_{th} is poorly constrained. The maximum value P_{th} could have is that required to maintain a completely liquid core in a well-mixed, adiabatic state¹³, for which $P_{th} = 4.7 \times 10^9 \text{ W}$ (Supplementary Table 1). It is likely that the correct value is smaller than this for two reasons: the presence of a growing inner core¹⁰ would reduce the power needed to maintain an adiabat; and under some conditions, through-going fluid motions can occur in a subadiabatic fluid if the outer boundary surface is ellipsoidal²¹. Therefore, our use of the adiabatic value as P_{th} is conservative, in that it will result in a minimum-duration estimation of the dynamo lifetime and underestimation of field intensity at a given Earth–Moon distance.

Figure 1 shows the evolution of the power dissipated at the CMB, P_Σ , compared with P_{th} . The power dissipated decreases rapidly with time as the semimajor axis increases and drops below the adiabatic threshold at $a \approx 48R_e$. In our nominal evolution model, this occurs 2.7 Gyr ago, but other temporal models would result in different threshold times (Supplementary Information, section 3), as would using different threshold values.

We estimate the magnetic field on the basis of the available power. This approach neglects the spatial pattern of the flow, which may be important^{16,22} and will require future work. There need be no simple connection between palaeointensity and available power because the dissipation can be on smaller scales than the large-scale current, but it is nonetheless useful to consider possible scalings for the field. For our first model (model A), we will derive a scaling for the lunar palaeofield based on Earth's current magnetic field (Methods):

$$B_{Am}(t) = B_e(t_n) \frac{R_e^3}{R_m^3} \left(\frac{R_{cm}}{R_{ce}}\right)^{5/2} \left(\frac{P_{dyn,m}(t)}{P_{dyn,e}(t_n)}\right)^{1/2} \quad (3)$$

$$\approx 1 \mu\text{T} \times \left(\frac{P_{dyn,m}(t)}{3 \times 10^{11} \text{ W}}\right)^{1/2}$$

Here R_{cm} and R_{ce} are the radii of the Moon and Earth's cores, R_m is the radius of the Moon, $P_{dyn,m}$ and $P_{dyn,e}$ are the respective powers available to drive a dynamo for the Moon and Earth, B_e is the terrestrial surface magnetic field strength and B_{Am} is the lunar surface magnetic field strength predicted by model A.

The numerical value in equation (3) was derived by using the values given in Supplementary Table 1. The largest source of uncertainty is in the value of the available power for Earth's core, $P_{dyn,e}$. Here we have taken $P_{dyn,e} = 10^{13} \text{ W}$ (ref. 13), which will result in a conservatively small lunar field strength (Supplementary Information, section 1).

For our second model (model B), we use a set of model results and direct scalings from ref. 23. We obtain

$$B_{Bm} \approx d \left(\frac{R_{cm}}{R_m}\right)^3 (2\mu_0 f_{ohm} \rho_c)^{1/2} \left(\frac{\alpha_c G}{3C_{p,c}}\right)^{1/3} P_{dyn,m}^{1/3} \quad (4)$$

$$\approx 6d \mu\text{T} \times \left[\frac{P_{dyn,m}(t)}{3 \times 10^{11} \text{ W}}\right]^{1/3}$$

where d is the ratio of the dipolar magnetic field to the total field at the CMB, μ_0 is the magnetic permeability in vacuum, f_{ohm} is the ratio of ohmic dissipation to total dissipation, ρ_c is the density of the core, α_c is the coefficient of linear thermal expansion of the core and $C_{p,c}$ is the heat capacity of the core (see Methods and Supplementary Table 1 for symbol values). Figure 2 shows how the surface field strengths based on our two different scalings (equations (3) and (4)) evolve with time. For model B, we show two curves, one with $d = 1$ and one with $d = 1/7$ (Methods and Supplementary Fig. 3). For both models, the predicted field strength decreases to zero once the available power is less than the threshold value ($a \approx 48R_e$, $t \approx 2.7$ Gyr ago). The field strengths decrease as the Moon moves outward, but at different rates. At 4 Gyr ago, the predicted field strengths are in the range $\sim 1\text{--}10 \mu\text{T}$. According to ref. 16, the surface field is expected to be dominantly dipolar and to undergo periodic reversals. However, because of predicted variations in pole orientation¹⁶ and likely impact-induced reorientation²⁴, existing inferences of magnetic palaeopole orientations based on remote sensing data⁴ do not provide a strong observational constraint on our model.

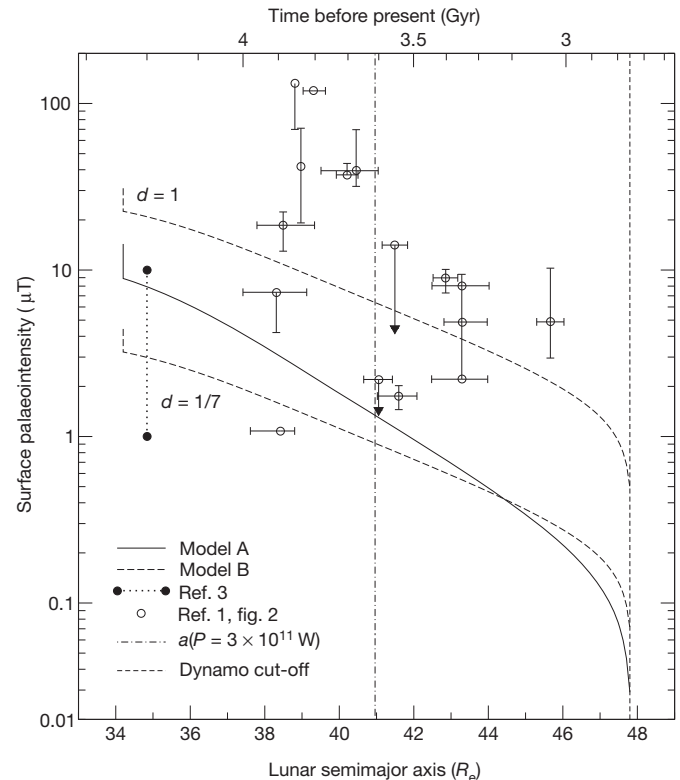


Figure 2 | The lunar surface magnetic palaeointensity predicted by our models. Model A (equation (3)) is shown by the solid line and model B (equation (4)) is shown by the (non-vertical) dashed lines; for model B, two lines are shown, with respective ratios of dipolar field strength to total field strength of $d = 1$ and $1/7$, as shown. The estimate range of ref. 3 is plotted as filled circles. Older palaeointensity data and error bars (from fig. 2 in ref. 1) are plotted as open circles. The conversion of semimajor axis to time is as in Fig. 1. The vertical dash-dot line indicates the distance at which the power dissipated is exactly $3 \times 10^{11} \text{ W}$, and the vertical dashed line denotes the minimum cut-off age of the dynamo (see text). For the purposes of display, the calculations were performed with no truncation of coefficients. The error bars are defined in ref. 1 as follows: “horizontal error bars reflect the published uncertainty values associated with the radiometric age determinations. Vertical error bars reflect variations in intensity between subsamples ... [and] uncertainties in the fit between the data and lines correspond to various paleofields”.

A recent estimate of lunar palaeointensity³ gives a value of $\sim 1\text{--}10\ \mu\text{T}$ at 4.2 Gyr ago (Fig. 2). This range is compatible with our models. Older palaeointensity estimates¹ for 3.0–4.0 Gyr ago are also plotted in Fig. 2, although these are probably less reliable^{2,12}. Many of the palaeointensities fall within our model range, although the high values corresponding to $\sim 3.6\text{--}3.9$ Gyr ago might require an additional driving mechanism^{11,25} or a smaller P_{th} . Nonetheless, our results suggest that a mechanically driven dynamo could persist for at least ~ 1.6 Gyr, which is much longer than any likely convection-driven dynamo.

Whereas the relationship between magnetic field intensity and lunar semimajor axis, a , is probably robust, the conversion of a to time is much more uncertain. For instance, although in our nominal model²⁰ $a = 48R_{\text{e}}$ occurs ~ 2.7 Gyr ago, in other models this distance might not occur until ~ 1.8 Gyr ago^{19,20,26,27} (Supplementary Information, section 3). Thus, our assessment of a billion-year lifetime for a mechanically stirred dynamo is probably conservative.

The simple relationship between dissipation and equatorial inclination shown in equation (1) is probably inappropriate at the earliest times. Stronger core–mantle coupling reduces the differential motion^{9,15,17}. Even when this coupling is weak, dissipation is probably self-limiting to $\sim 3 \times 10^{11}$ W because the overlying mantle will melt. At this dissipation rate, the core temperature increase over 0.1 Gyr (approximating the initial orbital evolution timescale) is about 1,000 K, which is enough to initiate mantle melting and reduce the effectiveness of core stirring.

Despite these caveats, however, our results raise several interesting possibilities. First, the model predicts a long-lived magnetic field that weakens with time, followed by an abrupt shut-off (Fig. 2). Thus, the palaeomagnetic record may ultimately be used to constrain the outward evolution of the Moon (and the dissipation rate within the Hadean Earth²⁸). Second, at the earliest distances ($a < 26R_{\text{e}}\text{--}29R_{\text{e}}$), when there is no differential motion¹⁵, our proposed mechanism does not generate a magnetic field. Therefore, we predict that rocks from this time period (according to our nominal model, this distance corresponds to before 4.45 Gyr ago) should be unmagnetized, unless other mechanisms were available^{11,25}. Third, similar mechanically driven dynamos may have operated in other bodies where tidally driven differential core–mantle motion may have occurred, such as large asteroids like 4 Vesta or the angrite parent body^{29,30}. Finally, our work also poses a challenge to numerical modellers of dynamos, namely that of determining the properties of mechanically driven dynamos.

METHODS SUMMARY

The relation between I_{e} and a . We fit the following explicit polynomial function (Supplementary Fig. 1) to the plot of obliquity versus semimajor axis, a (fig. 2 in ref. 18), using I_{e} = obliquity $\times 5.16^{\circ}$:

$$I_{\text{e}}(a) = \begin{cases} 71.84^{\circ}; a = 34.2R_{\text{e}} \\ 0.1075^{\circ} \times \tilde{a}^{10} - 0.0332^{\circ} \times \tilde{a}^9 - 1.0008^{\circ} \times \tilde{a}^8 \\ + 0.6110^{\circ} \times \tilde{a}^7 + 2.7016^{\circ} \times \tilde{a}^6 - 1.7281^{\circ} \times \tilde{a}^5 \\ - 2.3280^{\circ} \times \tilde{a}^4 - 1.4509^{\circ} \times \tilde{a}^3 + 6.9951^{\circ} \times \tilde{a}^2 \\ - 6.6208^{\circ} \times \tilde{a} + 5.5828^{\circ}; 34.2R_{\text{e}} \leq a \leq 60.2R_{\text{e}} \end{cases} \quad (5)$$

where $\tilde{a} = (a/R_{\text{e}} - 46.6308)/7.7288$.

Palaeointensity models. Full derivations of these models may be found in Methods.

In model A (equation (3)), we assume that we can, using appropriate parameters, scale Earth's magnetic field to the Moon. We assume that the magnetic field at the surface of the Moon can be approximated as a current loop in the core with the following characteristics: the radius of the current loop is proportional to the radius of the core and the cross-sectional area of the current loop is proportional to the squared radius of the core. We assume that Earth's field can be similarly characterized. This is not intended as anything more than a rough guide to what is possible. In reality, it is probable that much of the dissipation associated with Earth's field occurs on scales far smaller than the largest current loop that could fit within Earth's core. Nevertheless, it tells us what field to expect in the (unlikely) case of similarity of Earth and the Moon's dynamos.

In model B (equation (4)), rather than scaling from Earth we use a set of model results and direct scalings from ref. 23. We used their equation (2) to estimate the field intensity at the top of the CMB. We set the efficiency factor to be

$F \approx 4\pi G \mu_0 \rho_{\text{c}} R_{\text{cm}}^2 / 3C_{\text{p,c}}$, scaled the field from the CMB to the surface by including a factor of $(R_{\text{cm}}/R_{\text{m}})^3$, and scaled the dipolar to total field strength at the CMB using the parameter d .

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 May; accepted 15 September 2011.

- Cisowski, S. M., Collinson, D. W., Runcorn, S. K. & Stephenson, A. A review of lunar paleointensity data and implications for the origin of lunar magnetism. *J. Geophys. Res.* **88**, A691–A704 (1983).
- Wieczorek, M. A. *et al.* in *New Views of the Moon* (eds Jolliff, B. L., Wieczorek, M. A., Shearer, C. K. & Neal, C. R.) 60, 221–364 (Mineralogical Society of America, 2006).
- Garrick-Bethell, I., Weiss, B. P., Shuster, D. L. & Buz, J. Early lunar magnetism. *Science* **323**, 356–359 (2009).
- Hood, L. L. Central magnetic anomalies of Nectarian-aged lunar impact basins: probable evidence for an early core dynamo. *Icarus* **211**, 1109–1128 (2011).
- Stegman, D. R., Jellinek, A. M., Zatman, S. A., Baumgardner, J. R. & Richards, M. A. An early lunar core dynamo driven by thermochemical mantle convection. *Nature* **421**, 143–146 (2003).
- Takahashi, F. & Tsunakawa, H. Thermal core–mantle coupling in an early lunar dynamo: implications for a global magnetic field and magnetosphere of the early Moon. *Geophys. Res. Lett.* **36**, L24202 (2009).
- Konrad, W. & Spohn, T. Thermal history of the Moon: implications for an early core dynamo and post-accretional magmatism. *Adv. Space Res.* **19**, 1511–1521 (1997).
- Yoder, C. F. The free librations of a dissipative Moon. *Phil. Trans. R. Soc. Lond. A* **303**, 327–338 (1981).
- Williams, J. G., Boggs, D. H., Yoder, C. F., Ratcliff, J. T. & Dickey, J. O. Lunar rotational dissipation in solid body and molten core. *J. Geophys. Res.* **106**, 27933–27968 (2001).
- Weber, R. C., Lin, P.-Y., Garner, E. J., Williams, Q. & Lognonné, P. Seismic detection of the lunar core. *Science* **331**, 309–312 (2011).
- Hood, L. L. & Artemieva, N. A. Antipodal effects of lunar basin-forming impacts: initial 3D simulations and comparisons with observations. *Icarus* **193**, 485–502 (2008).
- Lawrence, K., Johnson, C., Tauxe, L. & Gee, J. Lunar paleointensity measurements: implications for lunar magnetic evolution. *Phys. Earth Planet. Inter.* **168**, 71–87 (2008).
- Nimmo, F. in *Core Dynamics* (ed. Olson, P.) 31–65 (Treatise on Geophysics 8, Academic, 2007).
- Goldreich, P. Precession of the Moon's core. *J. Geophys. Res.* **72**, 3135–3137 (1967).
- Meyer, J. & Wisdom, J. Precession of the lunar core. *Icarus* **211**, 921–924 (2011).
- Tilgner, A. Precession-driven dynamos. *Phys. Fluids* **17**, 034104 (2005).
- Peale, S. J. Generalized Cassini's laws. *Astron. J.* **74**, 483–489 (1968).
- Ward, W. R. Past orientation of the lunar spin axis. *Science* **189**, 377–379 (1975).
- Williams, G. E. Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit. *Rev. Geophys.* **38**, 37–59 (2000).
- Webb, D. J. Tides and the evolution of the Earth–Moon system. *Geophys. J. R. Astron. Soc.* **70**, 261–271 (1982).
- Cébron, D., Maubert, P. & Le Bars, M. Tidal instability in a rotating and differentially heated ellipsoidal shell. *Geophys. J. Int.* **182**, 1311–1318 (2010).
- Roberts, P. H., Glatzmaier, G. A. & Clune, T. L. Numerical simulation of a spherical dynamo excited by a flow of von Karman type. *Geophys. Astrophys. Fluid Dyn.* **104**, 207–220 (2010).
- Christensen, U. R., Holzwarth, V. & Reiners, A. Energy flux determines magnetic field strength of planets and stars. *Nature* **457**, 167–169 (2009).
- Ong, L. & Melosh, H. J. in *41st Lunar Planet. Sci. Conf. abstr.* 1363, (<http://www.lpi.usra.edu/meetings/lpsc2010/pdf/1363.pdf>) (Lunar and Planetary Institute, 2010).
- Le Bars, M., Wieczorek, M. A., Karatekin, Ö., Cébron, D. & Laneuville, M. An impact-driven dynamo for the early Moon. *Nature* (<http://dx.doi.org/10.1038/nature10565>) (this issue).
- Ooe, M., Sasaki, H. & Kinoshita, H. in *Variations in Earth Rotation* (eds McCarthy D. D. & Carter, W. E.) 51–57 (American Geophysical Union, 1990).
- Walker, J. C. G. *et al.* in *Earth's Earliest Biosphere* (ed. Schopf, J.) 260–290 (Princeton Univ. Press, 1983).
- Zahnle, K. *et al.* Emergence of a habitable planet. *Space Sci. Rev.* **129**, 35–78 (2007).
- Weiss, B. P. *et al.* Magnetism on the angrite parent body and the early differentiation of planetesimals. *Science* **322**, 713–716 (2008).
- Bills, B. G. & Nimmo, F. Forced obliquities and moments of inertia of Ceres and Vesta. *Icarus* **213**, 496–509 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements C.A.D. would like to thank B. P. Weiss and I. Garrick-Bethell for discussions.

Author Contributions D.J.S. thought of the initial idea. C.A.D. performed the calculations. F.N. investigated the conditions under which turbulence occurs. All authors discussed the results and implications and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.D. (cadwyer@ucsc.edu).

METHODS

The relation between I_c and a . The Moon is presently in a Cassini state and has been since at least¹⁸ the time at which $a = 34.2R_e$. Thus, there is a functional relationship between I_c and a over the range of a between that time and the present day. The relationship between I_c and a as determined by ref. 18 was numerical. We arrived at an explicit expression for I_c in the following way. We digitized the plot shown in fig. 2 of ref. 18 over the range of interest, converted obliquity to I_c (using $I_c = \text{obliquity} - 5.16^\circ$), smoothed the line and then fitted a tenth-order polynomial:

$$I_c(a) = \begin{cases} 71.84^\circ; a = 34.2R_e \\ 0.1075^\circ \times \tilde{a}^{10} - 0.0332^\circ \times \tilde{a}^9 - 1.0008^\circ \times \tilde{a}^8 \\ + 0.6110^\circ \times \tilde{a}^7 + 2.7016^\circ \times \tilde{a}^6 - 1.7281^\circ \times \tilde{a}^5 \\ - 2.3280^\circ \times \tilde{a}^4 - 1.4509^\circ \times \tilde{a}^3 + 6.9951^\circ \times \tilde{a}^2 \\ - 6.6208^\circ \times \tilde{a} + 5.5828^\circ; 34.2R_e \leq a \leq 60.2R_e \end{cases} \quad (5)$$

where $\tilde{a} = (a/R_e - 46.6308)/7.7288$. In Supplementary Fig. 1, we have plotted I_c and a against time before present (t), with the relationship between a and t calculated using the model described in the next section. Equation (5) is only an approximation to the relationship in ref. 18; however, given the great uncertainty in the a -to- t relation (Supplementary Information, section 3.1), the errors introduced by the use of equation (5) are not important. Likewise, errors in the result of ref. 18 are likely to be negligible in comparison with the uncertainties in the relation between a and t .

Converting a into time before present (t). As Cassini-state theory does not include time, a separate, independent relationship for the temporal evolution of the lunar orbit must be assumed. There is extreme uncertainty in this over the distances and times of interest, resulting in great temporal variability in any model (Supplementary Information, section 3). In addition to our nominal model (described below), we also examined four other possible models. We compare these models in Supplementary Information (section 3.2 therein and Supplementary Figs 2 and 3). We require that models for the lunar orbit fit the available geologic constraints¹⁹ and we assume that the Moon formed at about the Roche limit³¹ some time shortly after the Solar System formed.

The nominal model of this paper was modified from models c and d of ref. 20. Model c included only power dissipation from an ocean and model d included power dissipation in both the ocean and solid Earth. Neither model predicted lunar formation at ~ 4.6 Gyr ago, but the two predicted formation ages bracketed that value.

We digitized (a , t) points for both these models and used a weighted average to combine them into an evolution model in which the Moon formed 4.6 Gyr ago. The weighted mean was calculated in the following manner: the (a , t) curves were interpolated to have a common point spacing, and a weighted average of the t values at each a point was taken to derive the final (a , t) model. The weights were determined by applying the lever rule to the goal formation age (4.6 Gyr ago) and the curves' respective formation ages. This model is plotted in Supplementary Figs 1 and 2.

Palaeointensity models. We make the simplifying assumption that we may relate the surface magnetic field strength to the power available to the lunar dynamo. We consider two separate models for lunar palaeointensity: models A and B.

In model A (equation (3)), we assume that we can, using appropriate parameters, scale Earth's magnetic field to the Moon. We assume that the magnetic field at the surface of the Moon can be approximated as a current loop in the core with the following characteristics: the radius of the current loop is proportional to the radius of the core and the cross-sectional area of the current loop is proportional to the squared radius of the core. We assume that Earth's field can be similarly characterized. This is not intended as anything more than a rough guide to what is possible. In reality, it is probable that much of the dissipation associated with Earth's field occurs on scales far smaller than the largest current loop that could fit within Earth's core. Nevertheless, it tells us what field to expect in the (unlikely) case of similarity of Earth and the Moon's dynamos.

What follows is the derivation of the expression for the lunar surface palaeointensity as a function of available power (that is, equation (3) of the main text):

From electromagnetism it follows that

$$P = I^2 R_s$$

where P denotes power, I denotes current and R_s denotes resistance. Also, the intensity of a magnetic field is

$$B = \mu_0 I R_1^2 / 2D^3$$

where D is the distance between the centre of the current loop and the point where the field is measured and R_1 is the radius of the current loop. The resistance is

$$R_s = L / \sigma A_c$$

where σ denotes conductivity, L is the length of the current loop and A_c is the cross-sectional area of the current loop.

Combining the above, we have

$$B = \frac{\mu_0 R_1^2}{2D^3} \left(\frac{P \sigma A_c}{L} \right)^{1/2}$$

Because we are interested in the strengths of fields on the surface of the planet, $D = R_p$. Additionally, we assume that the following proportionalities hold: $L \propto R_c$, $R_1 \propto R_c$, $A_c \propto R_c^2$ (where R_c is the radius of the planet's core). This gives

$$B \propto \frac{P^{1/2} R_c^{5/2}}{R_p^3}$$

We can determine the constant of proportionality using Earth's present-day field.

Thus, with t_n denoting the present day, the palaeointensity of the lunar surface field as predicted by our model A is given by equation (3):

$$B_{Am}(t) = B_c(t_n) \frac{R_c}{R_m^3} \left(\frac{R_{cm}}{R_{ce}} \right)^{5/2} \left(\frac{P_{dyn,m}(t)}{P_{dyn,e}(t_n)} \right)^{1/2} \\ \approx 1 \mu T \times \left(\frac{P_{dyn,m}(t)}{3 \times 10^{11} W} \right)^{1/2}$$

In model B, rather than scaling from Earth we use a set of model results and direct scalings from ref. 23. According to equation (2) of ref. 23

$$\langle B \rangle^2 (2\mu_0)^{-1} = c f_{ohm} \langle \rho_c \rangle^{1/3} (F q_0)^{2/3}$$

Here B is the magnetic field strength within the dynamo; μ_0 is the permeability of free space; $q_0 = (P_\Sigma - P_{ad})(4\pi R_{cm}^2)^{-1}$ is the heat flux at the CMB in excess of the adiabatic value (see Supplementary Information of ref. 23); c , f_{ohm} and F are constants; and angle brackets denote averages over the volume over which the kinetic energy (from the differential rotation, in this case) is added to the core. The values of c and f_{ohm} given in Supplementary Table 1 are taken directly from ref. 23.

F is defined in equation (3) of ref. 23 as

$$F^{2/3} = \frac{1}{\text{Vol}} \int_{r_a}^{r_b} \left(\frac{q_c(r)}{q_0} \frac{L(r)}{H_T(r)} \right)^{2/3} \left(\frac{\rho(r)}{\langle \rho \rangle} \right)^{1/3} 4\pi r^2 dr \quad (6)$$

where Vol is the volume described above; r_a and r_b are respectively the lower and upper radial bounds of the volume; $q_c(r)$ is the power per unit area as a function of radius; $L(r)$ is the length scale of the largest convective structure; $H_T(r) = C_p(\alpha g(r))^{-1}$ is the temperature scale height as a function of radius; C_p is the heat capacity; α is the coefficient of linear thermal expansion; and $g(r)$ is the acceleration due to gravity as a function of radius.

Because the magnetic field strength depends on the total power and not on the spatial distribution with which it is deposited²³, we simplify this equation by assuming that all the terms within the integral are constant, and derive the following expression for F :

$$F \approx \frac{4\pi G \alpha_c \rho_c R_{cm}^2}{3 C_{p,c}} \quad (7)$$

Different assumptions (for example letting q_c vary linearly with r) yield results differing by a factor of order one.

We set the scale height of temperature to be $C_{p,c}(\alpha g_{cm})^{-1}$ and the scale height for convection to be R_{cm} , we set $g_{cm} = GM_{cm}/R_{cm}^2 = 4\pi G R_{cm} \rho_c / 3$, we take Vol to be the volume of the entire core and we approximate $\langle B \rangle$ by $B_m(R_{cm})$.

Because the lunar core has a small radius, only the dipolar part of the magnetic field is likely to be detectable at the surface. The magnetic field within the dynamo-generating region will contain higher-degree components^{16,23}. However, the ratio of the dipolar field strength to the total field strength at the CMB is somewhat uncertain, especially as precession-driven dynamos seem to have power spectra different from those of convective dynamos¹⁶. We have therefore defined a term, d , that is the ratio of the dipolar field strength to the total field strength at the top of the dynamo region (the CMB). For convection-driven dynamos²³, $d \approx 1/7$, and for the (probably unrealistic) purely dipolar case, $d = 1$. For $d < 1$, this correction has the result of reducing the predicted (dipolar) surface field strength (note that such a correction is not needed in model A). Taking this correction into account, the

lunar surface palaeointensity predicted by our model B is that given in equation (4):

$$B_{\text{Bm}} \approx d \left(\frac{R_{\text{cm}}}{R_{\text{m}}} \right)^3 (2\mu_0 c f_{\text{ohm}} \rho_c)^{1/2} \left(\frac{\alpha_c G}{3C_{p,c}} \right)^{1/3} P_{\text{dyn,m}}^{1/3} \\ \approx 6d \mu\text{T} \times \left[\frac{P_{\text{dyn,m}}(t)}{3 \times 10^{11} \text{ W}} \right]^{1/3}$$

Here the factor of $(R_{\text{cm}}/R_{\text{m}})^3$ scales the magnetic field strength from the CMB to the surface.

A further source of uncertainty in this model is in the calculation of the efficiency factor, F . Although this factor does depend on the spatial distribution of the power, it is the total power that really matters²³; thus, this uncertainty is unlikely to affect our conclusions significantly. An additional concern is the extent to which numerical models of terrestrial and stellar dynamos have the same scaling behaviour as mechanically driven dynamos of the kind we are proposing here (for example in the ratio of dipolar to higher-degree terms). We carry out a limited discussion of this topic in Supplementary Information, section 2, but further work will be needed.

31. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).

An impact-driven dynamo for the early Moon

M. Le Bars¹, M. A. Wieczorek², Ö. Karatekin³, D. Cébron¹ & M. Laneuville²

The origin of lunar magnetic anomalies^{1–5} remains unresolved after their discovery more than four decades ago. A commonly invoked hypothesis is that the Moon might once have possessed a thermally driven core dynamo³, but this theory is problematical given the small size of the core and the required surface magnetic field strengths⁶. An alternative hypothesis is that impact events might have amplified ambient fields near the antipodes of the largest basins⁷, but many magnetic anomalies exist that are not associated with basin antipodes. Here we propose a new model for magnetic field generation, in which dynamo action comes from impact-induced changes in the Moon's rotation rate. Basin-forming impact events are energetic enough to have unlocked the Moon from synchronous rotation⁸, and we demonstrate that the subsequent large-scale fluid flows in the core, excited by the tidal distortion of the core–mantle boundary⁹, could have powered a lunar dynamo. Predicted surface magnetic field strengths are on the order of several microteslas, consistent with palaeomagnetic measurements⁵, and the duration of these fields is sufficient to explain the central magnetic anomalies associated with several large impact basins.

Magnetic field measurements of the Moon from orbit demonstrate that portions of its crust are strongly magnetized¹, and palaeomagnetic analyses of lunar rocks show that some samples possess stable remanent magnetizations². However, the origin of the fields that magnetized the crust and the manner by which the surface field strength varied with time are still being debated^{3–5}. Two facts lead us to propose a new model for lunar magnetic field generation, in which a global field is generated by a core dynamo that is powered by changes in the Moon's rotation rate following large impact events. The first observation is that six large impact basins of Nectarian age (which are more than about 4 billion years old) possess central magnetic anomalies^{10,11} (Fig. 1 and Supplementary Fig. 1). These basins represent the clearest examples of areas in which endogenous lunar features correlate with magnetic anomalies, and for that reason offer an invaluable clue to the origin of lunar magnetism. It is improbable that these magnetic anomalies are caused by the mare basaltic lava flows present in these basins, because the mare basalts are only weakly magnetic² and similarly aged basalts elsewhere on the Moon do not possess magnetic anomalies. Impact cratering studies predict these events to have melted substantial quantities of both the crust and the projectile, forming impact melt sheets several kilometres thick¹². Lunar impact melt rocks are by far the most magnetic igneous rocks on the Moon², and the central magnetic anomalies associated with the Nectarian-age basins most plausibly formed as their melt sheets cooled through their Curie temperatures in the presence of a magnetic field. Given the large conductive cooling timescales of these thick deposits, a stable magnetic field must have been present for thousands of years immediately following the impact event (Supplementary Information, section 4, and Supplementary Fig. 10), and the only plausible means of generating such a field is a core dynamo.

The second fact that bears on the origin of impact-basin-related magnetic anomalies is that each of the corresponding impact events would have significantly affected the rotational state of the Moon,

either by unlocking it from synchronous rotation (Supplementary Information, section 3.1, and Supplementary Fig. 3) or by setting up large-amplitude free librations⁸ (Supplementary Information, section 3.2). The rotation rate of the solid portion of the Moon would have been instantaneously modified following these impacts, but its molten

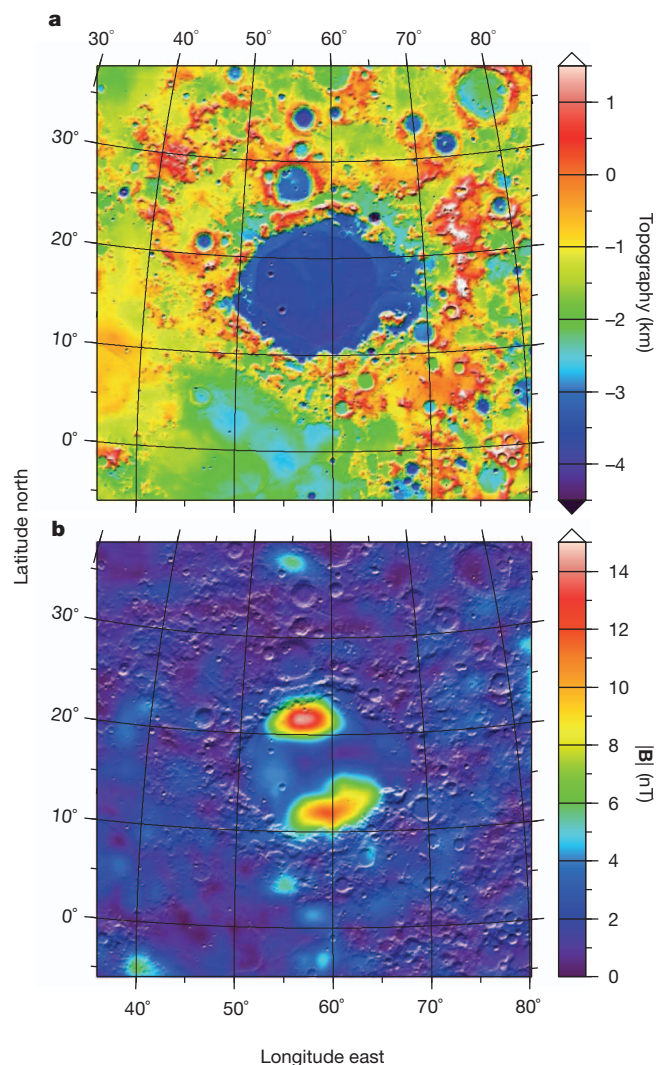


Figure 1 | Surface topography and total magnetic field strength of the ~550-km-diameter Crisium impact basin. **a**, Topography; **b**, magnetic field strength ($|B|$). Two prominent anomalies are confined to the interior of this basin, where a thick impact melt sheet is predicted. The magnetic field strength is derived from the global sequential Lunar-Prospector-based model¹ evaluated 30 km above the mean planetary radius, and the topography is from the Lunar Reconnaissance Orbiter laser altimeter data³⁰. Both images are displayed in a Lambert azimuthal equal-area projection with an image width and height of 1,200 km, and are overlain by a shaded relief map derived from the surface topography.

¹IRPHE, CNRS and Aix-Marseille Université, 49 rue F. Joliot Curie, BP 146, 13384 Marseille Cedex 13, France. ²Institut de Physique du Globe de Paris, Université Paris Diderot, 4 Avenue de Neptune, 94100 Saint Maur des Fossés, France. ³Royal Observatory of Belgium, 3 Avenue Circulaire, BE 1180 Uccle, Belgium.

core¹³ would have initially kept rotating at the pre-impact synchronous rate and would then have accelerated or decelerated much more gradually. Differential rotation at the core–mantle boundary would persist for thousands of years, potentially giving rise to hydrodynamic instabilities that could drive a core dynamo. Not only would these basin-forming impacts set up conditions under which highly magnetic impact melt rocks could acquire a thermoremanent magnetization, but they could also power a core dynamo to magnetize these rocks.

A huge amount of energy is stored in the spin and orbital motions of any planet, and under certain circumstances a portion of this energy can be used to drive three-dimensional flows in a fluid core. One mechanism for this is the excitation of hydrodynamic instabilities known as inertial instabilities. These instabilities involve a resonance between two inertial waves of the rotating fluid and a large-scale natural forcing, such as that from precession^{14,15}, librations¹⁶ or tides⁹ (Supplementary Information, section 1). The resulting fluid flows are highly energetic (see the demonstration in ref. 15 in response to the initial criticism in refs 17, 18) and dynamo capable, as has already been demonstrated for precession^{19,20} (Supplementary Information, section 2). Regarding instabilities excited by tides (tidal instability²¹), laboratory experiments⁹ have shown that a fully three-dimensional turbulent flow develops in the bulk fluid when the ratio between the equatorial ellipticity of the core–mantle boundary, β , and the square root of the Ekman number, E (which represents the ratio between viscous and Coriolis forces), is larger than a critical value of order one, and when a difference in angular velocity exists between the mean rotation of the fluid and the elliptical distortion of the fluid container. The root mean squared velocity of the flow is then of the same order of magnitude as the differential rotation²². In the case of the Moon, a lower-bound estimate (Supplementary Information, section 3, and Supplementary Fig. 2) gives values of β between 1.9×10^{-5} today and 1.5×10^{-4} when the Earth–Moon separation was about half of its current value, with corresponding Ekman numbers from 3×10^{-12} to 10^{-12} . The ratio $\beta/E^{1/2}$ is always greater than ten, allowing tidal instabilities to develop in the lunar core over its entire history, provided that an instantaneous non-zero differential rotation is imposed between the fluid core and mantle ellipticity.

The impact events that formed the large Nectarian-age impact basins with magnetic anomalies could all have unlocked the Moon from synchronous rotation⁸ (Supplementary Information, section 3.1). Given the relatively short resynchronization timescales, the pre-impact tidal deformation of the core–mantle boundary would remain frozen in the mantle, thus providing a differential rotation, $\Delta\Omega$, between the fluid core and the elliptical distortion of the core–mantle boundary. To generate a dynamo, two additional criteria must be met (Methods and Supplementary Information, section 2): the time it takes for the tidal instability to establish a fully turbulent state must be shorter than the time it takes the core to spin up or down to the mantle rotation rate, and the core flow must be vigorous enough that the magnetic Reynolds number is larger than the threshold value for dynamo action. The characteristic amplitude of the magnetic field intensity in the core is then estimated by adapting scaling relations developed for convective dynamos^{23–25} to our case of mechanical forcing. As for convective dynamos, the magnetic field strength at saturation is controlled by the available mechanical power rather than by any force balance. By assuming a turbulent core–mantle boundary layer^{13,26} as well as a rapid growth of the dynamo process, the magnetic field strength at the surface of the Moon is estimated to be (Methods)

$$B \approx f[0.0026\mu_0\rho R^2(\Delta\Omega)^2]^{1/2}(R/R_{\text{Moon}})^3 \quad (1)$$

where R_{Moon} and R are respectively the mean radii of the Moon and its core, ρ is the core density, μ_0 is the magnetic constant (the permeability of free space) and f is a prefactor of order one that expresses both the efficiency of power conversion from viscous dissipation to Joule dissipation and the ratio of the dipolar component of the exterior magnetic

field strength to the total core field strength. As a conservative estimate, we use $f = 0.13$ (Methods).

The expected magnetic field strength at the surface of the Moon is shown in Fig. 2 for Earth–Moon separations ranging between 25 and 55 Earth radii and for post-impact spin periods of between 3 and 35 days (see also the more detailed discussion in Supplementary Information, section 3.1, and Supplementary Figs 4–7). The Earth–Moon separation was probably less than 45 Earth radii in the Nectarian period, and the current separation is 60 Earth radii. The post-impact rotation rates following the formation of a 700-km-diameter basin with an average impact geometry are shown in red in Fig. 2 (Supplementary Information, section 3.1, and Supplementary Fig. 3). Other impact geometries, or a succession of impacts closely spaced in time, could have given rise to either larger or smaller changes in the rotation rate. In Fig. 2a, the tidal instability growth rates were calculated using the hydrostatic core–mantle boundary ellipticity. Given that the lithosphere of the Moon is not in hydrostatic equilibrium, this generates an additional gravitational potential at the core–mantle

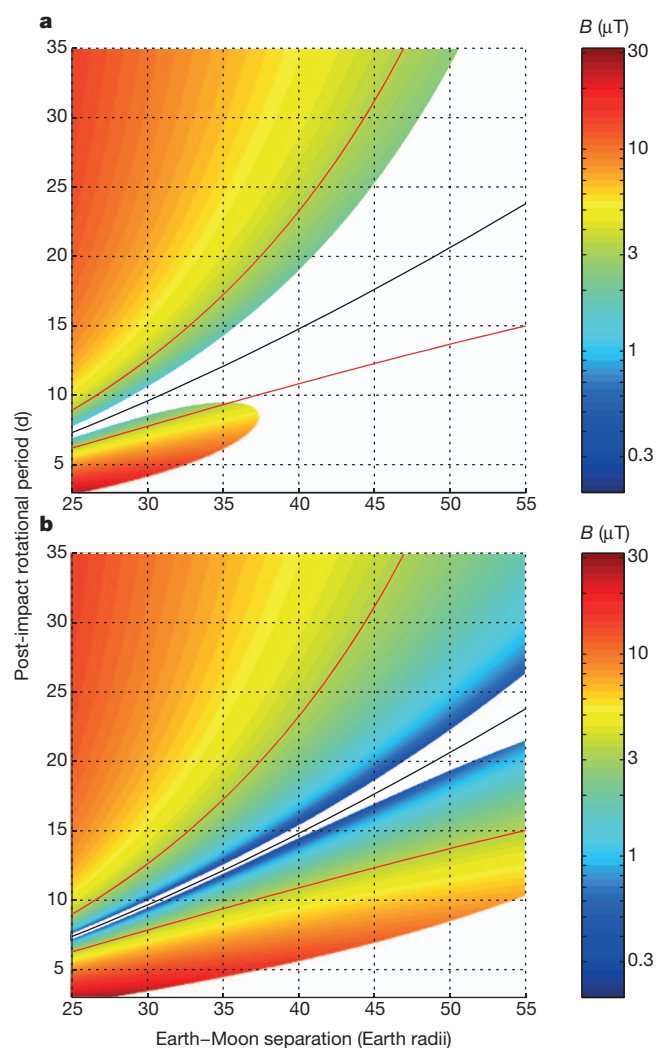


Figure 2 | Estimated magnetic field strength at the Moon's surface as a function of Earth–Moon separation and post-impact rotational period. Field strengths are plotted only when dynamo generation is possible. Also shown are the synchronous rotation rate of the Moon (black lines) and the expected range of rotational periods following the formation of a 700-km-diameter basin by a 5 km s^{-1} impact under average impact conditions (bounded by red lines; Supplementary Information, section 3.1, and Supplementary Fig. 3). The core–mantle boundary ellipticity is assumed to be equal to that of a purely hydrostatic Moon in **a**, and is assumed to be ten times this value in **b**.

boundary, causing the ellipticity to be greater than that of a purely hydrostatic Moon²⁷. As an extreme case, the magnetic field strength is plotted in Fig. 2b using an ellipticity ten times larger than the purely hydrostatic value (Supplementary Information, section 3). Although the core–mantle boundary ellipticity does not affect the amplitude of the generated magnetic field at saturation, it does influence the growth rates of the tidal instability and, hence, the parameter space in which it can develop.

Dynamo action driven by tidal instabilities is possible over a large range of post-impact rotation periods and Earth–Moon separations. Predicted magnetic field strengths are significant, and range from about 0.2 to 4 μT at the surface of the Moon following the formation of a 700-km-diameter basin. These values compare favourably to those obtained from lunar palaeomagnetic analyses⁵ that imply minimum field strengths of $\sim 1 \mu\text{T}$. Given the uncertainty of the prefactor, f , in equation (1), the actual field strengths could be up to about ten times larger. On the basis of the core spin-up times, the duration for each impact-induced dynamo is predicted to be between about 2×10^3 and 8×10^3 years, which would allow about 1 km of impact melt to cool through its Curie temperature and acquire a thermoremanent magnetization (Fig. 3). At the point where the Moon becomes synchronously locked, it would go through a phase of large-amplitude free librations, and these librations could also power a second dynamo, generating surface field strengths on the order of 1 μT (Supplementary Information, section 3.2, and Supplementary Figs 8 and 9).

Apart from forming magnetic anomalies associated with the impact melt sheets of large basins, it is also possible that impact-induced dynamos could have simultaneously magnetized portions of the proximal and distal ejecta of the basin. Magmatic intrusions and extrusive volcanic rocks that were cooling through their Curie temperatures when these dynamos were operating would have acquired a thermoremanent magnetization. Smaller impact events during this time could have magnetized crustal rocks elsewhere by the process of shock remanent magnetization²⁸. Furthermore, in addition to the six Nectarian-age impact craters that possess central magnetic anomalies, other impacts could have unlocked the Moon from synchronous rotation⁸, potentially powering dynamos in the Imbrian and pre-Nectarian periods. The lack of a clear central magnetic anomaly with these impact basins could be explained by either post-impact processes or differences in the abundance of magnetic carriers in the impact melt sheet.

Dynamos powered by impact-induced changes in rotation rate and subsequent tidal instabilities could thus explain a large portion of lunar

magnetic anomalies and natural remanent magnetizations of lunar samples. Previous studies have also suggested the possibility of generating dynamos by tidal instabilities on Io¹⁶ and Mars²⁹, and the formalism presented in this work makes it possible not only to test these hypotheses quantitatively, but to estimate the magnitude of the surface magnetic field strengths. Similar time-variable tidal deformations on Mercury, Ganymede, (early) Earth and some exoplanets could potentially account for various aspects of the magnetic fields observed with these bodies.

METHODS SUMMARY

The saturation strength of the magnetic field given by equation (1) is estimated following previous works on convective dynamos. We first suppose that the ratio between the magnetic energy and the Joule dissipation is proportional to the magnetic dissipation time²³. Following ref. 24, we then relate the Joule dissipation to the power dissipated by the elliptical instability. The mechanical power is evaluated following refs 9 and 22 by assuming that it is equal to the viscous dissipation at the Ekman boundary layer adjacent to the core–mantle boundary. This boundary layer results from the differing velocities of the fluid core and the solid mantle. This approach has been extended to account for the turbulent boundary layer expected for the Moon, following ref. 26. In plotting Fig. 2, the growth time of the tidal instability is determined for various Earth–Moon separations and spin periods of the mantle after impact using the analytical formula given in ref. 9. Also, we estimate the magnetic Reynolds number of the flow using the differential velocity between the mantle and the core. The magnetic field amplitude given by equation (1) is plotted when the magnetic Reynolds number is above the threshold for dynamo generation (estimated to be 1,000; Supplementary Information, section 2) and when the growth time of the elliptical instability is less than the core spin-up time. In Fig. 3, the time evolution of the differential rotation between the core and the mantle is calculated using a coupled numerical simulation that takes into consideration tidal and friction torques with the parameters in Supplementary Table 1. The field strength is estimated using the stationary equation (1), assuming a quasistatic evolution of a saturated magnetic field. The depth at which the Curie temperature is reached is estimated by the conductive cooling of a semi-infinite half-space.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 June; accepted 15 September 2011.

- Purucker, M. E. & Nicholas, J. B. Global spherical harmonic models of the internal magnetic field strength of the Moon based on sequential and coestimation approaches. *J. Geophys. Res.* **115**, E12007 (2010).
- Fuller, M. & Cisowski, S. M. in *Geomagnetism* (ed. Jacobs, J. A.) 307–455 (Academic, 1987).
- Stegman, D. R., Jellinek, A. M., Zatman, S. A., Baumgardner, J. R. & Richards, M. A. An early lunar core dynamo driven by thermochemical mantle convection. *Nature* **421**, 143–146 (2003).
- Lawrence, K., Johnson, C., Tauxe, L. & Gee, J. Lunar paleointensity measurements: implications for lunar magnetic evolution. *Phys. Earth Planet. Inter.* **168**, 71–87 (2008).
- Garrick-Bethell, I., Weiss, B. P., Shuster, D. L. & Buz, J. Early lunar magnetism. *Science* **323**, 356–359 (2009).
- Wieczorek, M. A. *et al.* The constitution and structure of the lunar interior. *Rev. Mineral. Geochem.* **60**, 221–364 (2006).
- Hood, L. L. & Artemieva, N. A. Antipodal effects of lunar basin-forming impacts: initial 3D simulations and comparisons with observations. *Icarus* **193**, 485–502 (2008).
- Wieczorek, M. A. & Le Feuvre, M. Did a large impact reorient the Moon? *Icarus* **200**, 358–366 (2009).
- Le Bars, M., Lacaze, L., Le Dizès, S., Le Gal, P. & Rieutord, M. Tidal instability in stellar and planetary binary systems. *Phys. Earth Planet. Inter.* **178**, 48–55 (2010).
- Halekas, J. S., Lin, R. P. & Mitchell, D. L. Magnetic fields of lunar multi-ring impact basins. *Meteorit. Planet. Sci.* **38**, 565–578 (2003).
- Hood, L. L. Central magnetic anomalies of Nectarian-aged lunar impact basins: Probable evidence for an early core dynamo. *Icarus* **211**, 1109–1128 (2011).
- Cintala, M. J. & Grieve, R. A. F. Scaling impact melting and crater dimensions: implications for the lunar cratering record. *Meteorit. Planet. Sci.* **33**, 889–912 (1998).
- Williams, J. G., Boggs, D. H., Yoder, C. F., Ratcliff, J. T. & Dickey, J. O. Lunar rotational dissipation in solid body and molten core. *J. Geophys. Res.* **106**, 27933–27968 (2001).
- Malkus, W. V. R. Precession of the Earth as the cause of geomagnetism. *Science* **160**, 259–264 (1968).
- Kerswell, R. R. Upper bounds on the energy dissipation in turbulent precession. *J. Fluid Mech.* **321**, 335–370 (1996).

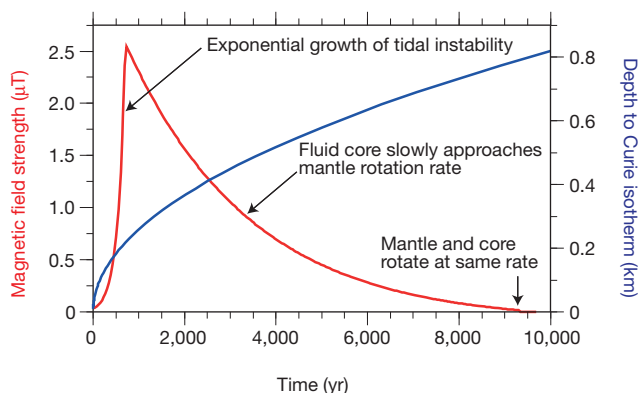


Figure 3 | Time evolution of the surface magnetic field strength (red) and depth to the Curie temperature in an impact melt sheet (blue) following an impact event. Here we consider an impact event that formed a 700-km-diameter basin at an Earth–Moon separation of 45 Earth radii. The spin period of the mantle following the impact is 31.2 days. The field strength in the exponential growth phase was estimated by assuming initial velocity perturbations in the core equal to 1% of the differential rotation and by using a core–mantle boundary ellipticity five times larger than the purely hydrostatic value.

16. Kerswell, R. R. & Malkus, W. V. R. Tidal instability as the source for Io's magnetic signature. *Geophys. Res. Lett.* **25**, 603–606 (1998).
17. Rochester, M. G., Jacobs, J. A., Smylie, D. E. & Chong, K. F. Can precession power the geomagnetic dynamo? *Geophys. J. R. Astron. Soc.* **43**, 661–678 (1975).
18. Loper, D. E. Torque balance and energy budget for the precessionally driven dynamo. *Phys. Earth Planet. Inter.* **11**, 43–60 (1975).
19. Tilgner, A. Precession driven dynamo. *Phys. Fluids* **17**, 034104 (2005).
20. Wu, C.-C. & Roberts, P. H. On a dynamo driven by topographic precession. *Geophys. Astrophys. Fluid Dyn.* **103**, 467–501 (2009).
21. Kerswell, R. R. Elliptical instability. *Annu. Rev. Fluid Mech.* **34**, 83–113 (2002).
22. Cébron, D., Le Bars, M., Leontini, J., Maubert, P. & Le Gal, P. A systematic numerical study of the tidal instability in a rotating triaxial ellipsoid. *Phys. Earth Planet. Inter.* **182**, 119–128 (2010).
23. Christensen, U. R. & Tilgner, A. Power requirement of the geodynamo from ohmic losses in numerical and laboratory dynamos. *Nature* **429**, 169–171 (2004).
24. Christensen, U. R. & Aubert, J. Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. *Geophys. J. Int.* **166**, 97–114 (2006).
25. Christensen, U. R., Holzwarth, V. & Reiners, A. Energy flux determines magnetic field strength of planets and stars. *Nature* **457**, 167–169 (2009).
26. Yoder, F. & Hutchison, R. The free librations of a dissipative Moon. *Phil. Trans. R. Soc. Lond. A* **303**, 327–338 (1981).
27. Meyer, J. & Wisdom, J. Precession of the lunar core. *Icarus* **211**, 921–924 (2011).
28. Gattacceca, J. *et al.* Unraveling the simultaneous shock magnetization and demagnetization of rocks. *Earth Planet. Sci. Lett.* **299**, 42–53 (2010).
29. Arkani-Hamed, J. Did tidal deformation power the core dynamo of Mars? *Icarus* **201**, 31–43 (2009).
30. Smith, D. E. *et al.* Initial observations from the Lunar Orbiter Laser Altimeter (LOLA). *Geophys. Res. Lett.* **37**, L18204 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements M.L.B. and D.C. acknowledge discussions with P. Le Gal and S. Le Dizès. Ö.K. acknowledges the support of the Belgian PRODEX programme.

Author Contributions M.L.B. and D.C. performed the fluid mechanics analysis, M.A.W. contributed to the analysis of the basin magnetic anomalies and post-impact rotation rates, Ö.K. contributed to analysis of the post-impact rotational evolution of the Moon and M.L. performed the impact melt thermal evolution analysis. All authors contributed to the conclusions presented in the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.L.B. (lebars@irphe.univ-mrs.fr).

METHODS

Saturation strength of the magnetic field. The saturation strength of the magnetic field, B , generated by a tidal dynamo, as given by equation (1), can be estimated by adapting scaling relationships previously derived for convective dynamos to our case of mechanical forcing. Similar to refs 24, 25, we assume that the magnetic field strength in the core is controlled by the available mechanical power rather than by any force balance. For the elliptical instability considered here, we know that thermal effects to first order do not influence the amplitudes of the tidally generated flow pattern^{31,32}. We thus model the core as a homogeneous, incompressible, electrically conductive fluid and neglect any thermal or stratification effects. Because precessional (and presumably tidal) dynamos do not depend on a driving thermal convection, we assume that all mechanical power that is dissipated by the instability is available for dynamo generation.

The mechanical power dissipated by elliptical instabilities has been evaluated in ref. 9 and validated numerically in ref. 22 (see equation (19) and fig. 10 therein) for the simplest form of the instability, the ‘spin-over’ mode, which corresponds to a rigid-body rotation of the fluid along an inclined axis with respect to the spin axis. As given by equation (5) in ref. 9, this mechanical power scales as

$$P \approx \frac{8}{3} \pi R^3 \eta \delta \Omega^2 \frac{R}{\delta} \quad (2)$$

where $\eta = \nu \rho$ is the core dynamic viscosity (ν is the core kinematic viscosity and ρ is the core density) and δ is the Ekman boundary layer thickness, $\delta \approx R \sqrt{E}$ (see, for example, ref. 33), with an Ekman number, $E = \nu / (\Omega_{\text{mantle}} R^2)$, that depends on Ω_{mantle} , the mean spin rate of the mantle. This viscous dissipation results from the differing velocities of the fluid core and the solid mantle at the laminar Ekman boundary layer adjacent to the core–mantle boundary. By neglecting the bulk turbulence that would develop in the fluid interior at the low Ekman numbers relevant for planetary cores, equation (2) gives a lower-bound estimate for the dissipated mechanical power. The dissipation of energy associated with elliptical instabilities that are more complicated than the spin-over mode will also occur mainly by viscous dissipation of kinetic energy through the Ekman layer. This means that the scaling powers in the various parameters of equation (2) should not depend strongly on the selected mode. Nevertheless, we acknowledge that a prefactor f_p , most probably larger than one, should be included in equation (2) to account for the small-scale motions associated with the more complicated instabilities.

The same expression as equation (2) is derived for the mechanical power dissipated by the precession of the lunar mantle in ref. 13 (see equations (81a) and (54) therein). This is not surprising given the similarity between the spin-over mode of the elliptical instability and the so-called Poincaré tilt-over mode excited by precession³⁴. Nevertheless, it is important to emphasize that the elliptical instabilities of interest here are expected to induce motions with much larger amplitudes than the small precessional flow considered in ref. 13, where $\Delta\Omega = \Omega_{\text{mantle}} \sin(I)$ with I being the (small) angle between the equatorial and ecliptic planes. Even in the limit of small-amplitude flows, a more reasonable estimate of the Moon’s core–mantle friction, based on the onset of fluid turbulence in the boundary layer, should be considered²⁶: in this case, the rate of dissipation depends on an ‘eddy’ viscosity rather than on the molecular viscosity. The problem of turbulent core–mantle coupling has still to be resolved explicitly, but a reasonable estimate is given in ref. 26 and redeveloped in ref. 13 (see equations (81a) and (55) in ref. 13), leading to

$$P \approx f_p \frac{3}{4} \pi^2 \kappa R^5 \rho |\Delta\Omega|^3 \quad (3)$$

where κ is a constant of the order of 7.3×10^{-4} for a 350-km-radius core with a viscosity of $10^{-6} \text{ m}^2 \text{ s}^{-1}$. As introduced above, f_p is a constant of order one that is related to the complexity of the selected mode of the elliptical instability.

To estimate the magnetic field strength at saturation associated with an elliptical instability, we follow ref. 24 and relate the power, P , dissipated by the elliptical instability to the Joule dissipation, D_{ohm} , through $D_{\text{ohm}} = f_{\text{ohm}} P$, where f_{ohm} is a constant less than one that depends on the energy source powering the dynamo. Further supposing that the ratio between the magnetic energy

$$E_B = \frac{4}{3} \pi R^3 \frac{B_{\text{core}}^2}{2\mu_0}$$

and the Joule dissipation is proportional to the magnetic dissipation time (ref. 24)

$$\frac{E_B}{D_{\text{ohm}}} \propto \frac{R}{u}$$

where u is the characteristic velocity driven by the instability ($u \approx R\Delta\Omega$ for elliptical instabilities²²), we find characteristic magnetic field strengths in the core of

$$B_{\text{core,L}} \approx \sqrt{4f_{\text{ohm}} f_p \mu_0 \rho R |\Delta\Omega| \sqrt{\nu \Omega_{\text{mantle}}}} \quad (4)$$

for the laminar estimate of power dissipation (equation (2)) and

$$B_{\text{core,T}} \approx \sqrt{\left(\frac{9\pi}{8}\right) f_{\text{ohm}} f_p \kappa \mu_0 \rho R^2 (\Delta\Omega)^2} \quad (5)$$

for the turbulent estimate of power dissipation (equation (3)). Because of the attenuation of magnetic anomalies with height, it is likely that only the dipolar component of the magnetic field will be important at the surface, which leads to a magnetic field estimate at the planet surface of

$$B \approx f_{\text{dip}} B_{\text{core}} (R/R_{\text{planet}})^3 \quad (6)$$

where R_{planet} is the mean planetary radius and f_{dip} is the ratio of the dipolar component of the magnetic field strength just outside the core to the total magnetic field strength inside the core–mantle boundary. Equation (1) directly derives from the combination of equations (5) and (6). We note that the value of the ellipticity of the core–mantle boundary, β , does not appear in these formulae: the value of β determines whether or not the elliptical instability can be excited, but once the flow has developed the induced motions and magnetic field depend only on the differential rotation between the fluid core and the mantle.

Equations (4), (5) and (6) represent the most up-to-date estimate of the magnetic field strength driven by an elliptical instability, but several sources of uncertainty exist in this estimate, mainly as a result of the unknown prefactor, $f_{\text{dip}} \sqrt{f_{\text{ohm}} f_p}$. The efficiency factor for the conversion of mechanical power to Joule dissipation, f_{ohm} , has been estimated²⁴ to range between about 0.2 and 0.8 in a numerical model of convective dynamos, but it has been argued²⁵ that this should be closer to 0.88 for the Earth. Moreover, on the basis of magnetohydrodynamic simulations²⁴, and as further discussed in ref. 25, the ratio between the external dipolar magnetic field strength and the internal core field strength, f_{dip} , is about 1/7 for Earth-like convective dynamos. However, we suspect that the prefactor f_p , which is equal to one for the simple spin-over mode, could be greater than one for more-complicated flow patterns. In addition to these prefactors, the core radius of the Moon is uncertain, and a ± 50 -km uncertainty would affect the surface magnetic field strength by a factor of about 1.6. An uncertainty of a factor of ten in the core molecular viscosity would yield an uncertainty of a factor of about 1.7 in the total field strength estimated by the laminar scaling law. And, finally, as discussed in ref. 13, various estimates for the turbulent coupling parameter, κ , lie in the range 0.0005–0.0021: with respect to our nominal value of 7.3×10^{-4} , the magnetic field strength could be uncertain by a factor of about 1.7. Considering the above uncertainties, but also claiming that the relevant physics are taken into account in the above derivation, we consider equations (4) and (5) to be precise to within a factor of order one. To provide a reliable lower-bound estimate for the surface magnetic field strength of the Moon from equation (6), we use a combined prefactor of $f_{\text{dip}} \sqrt{f_{\text{ohm}} f_p} = 0.13$ in our magnetic field calculations, with $f_{\text{dip}} = 1/7$, $f_{\text{ohm}} = 0.88$ and the restrictive value $f_p = 1$.

Growth rate of the tidal instability. After a large impact event desynchronizes the Moon’s rotation, a differential rotation, $\Delta\Omega$, will exist between the fluid core and the elliptically deformed core–mantle boundary. The pre-impact elliptical shape of core–mantle boundary is assumed to remain ‘frozen’ into the mantle during the short time it takes the Moon to become resynchronized by tidal forces. The differential rotation can be either positive or negative depending on whether the impact accelerates or decelerates the rotation rate of the Moon, and will persist as long as the fluid core has not spun up (or spun down) to the rotation rate of the mantle. When a rotating spherical container and enclosed fluid are initially in rigid-body rotation, and when the container is suddenly set spinning with a different angular velocity, Ω_{mantle} , the mean angular velocity of the fluid will spin up or down exponentially quickly to this new rotation rate with a characteristic timescale given by³³

$$T_{\text{spin-up}} = \frac{1}{\Omega_{\text{mantle}} \sqrt{E}} \quad (7)$$

This core spin-up time should be considered approximate, as it assumes laminar flow. The development of an elliptical instability would give rise to three-dimensional fluid flows that could alter this value. The rotation rate of the mantle will progressively resynchronize with the orbital motion on a timescale given by³⁵

$$T_{\text{sync}} = \frac{2\Omega_{\text{mantle}} d^6 C}{3GM_{\text{Earth}}^2 R_{\text{Moon}}^5 k_2} \frac{Q}{Q}$$

where d is the Earth–Moon separation, C is the lunar polar moment of inertia, Q is the quality factor, G is the gravitational constant, M_{Earth} is the mass of the Earth and k_2 is the degree-2 Love number. This timescale is considerably longer than the core spin-up timescale given by equation (7). Assuming a quasisteady state, where the

core spin-up time is greater than the growth time of the elliptical instability, the growth rate of the tide-driven elliptical instability can be analytically determined from the results of ref. 9:

$$\sigma_{\text{growth}} = \frac{(3\Delta\Omega + 2\Omega_{\text{mantle}})^2}{16(\Delta\Omega + \Omega_{\text{mantle}})^2} \beta |\Delta\Omega| - \alpha \sqrt{E} \Omega_{\text{mantle}} \quad (8)$$

Here α is a constant of order one that accounts for the amplitude of the viscous dissipation over the core–mantle boundary and which depends on the inertial waves that are excited. In the following, we take $\alpha = 2.62$, which is the relevant value for the spin-over mode³⁶. The characteristic growth time is then simply given by $T_{\text{growth}} = 1/\sigma_{\text{growth}}$. We note that, as mentioned in ref. 9, instabilities cannot occur in a ‘forbidden zone’, $-3/2 < \Omega_{\text{mantle}}/\Delta\Omega < -1/2$, where no resonance involving inertial waves and tidal forcing is possible. We also note that T_{growth} is actually a lower-bound estimate for the time necessary for the turbulent flow driven by the elliptical instability to be established, because it supposes that the instability starts from a stable laminar two-dimensional rotating flow. If the core were already unstable before the impact (for instance because the effects of a previous impact have not yet been fully dissipated), the elliptical instability would saturate and generate a fully three-dimensional flow over a much shorter period than the typical growth time.

In plotting Fig. 2, we determine the growth time of the tidal instability given by equation (8) for various Earth–Moon separations and spin periods of the mantle after impact. We also estimate the magnetic Reynolds number of the flow using the differential velocity between the mantle and the core. The magnetic field amplitude given by equation (1) is plotted when the magnetic Reynolds number is above the threshold for dynamo generation (estimated to be 1,000; Supplementary Information, section 2) and when the growth time of the elliptical instability is

less than the core spin-up time (equation (7)). In Fig. 3, the time evolution of the differential rotation between the core and the mantle is calculated using a coupled numerical simulation by considering core–mantle friction as well as inelastic deformations of the Moon due to the gravitational effect of Earth (see details in ref. 37). We applied a backward finite-difference scheme with the parameters in Supplementary Table 1 and the present-day k_2/Q ratio for the Moon. The core and mantle are assumed to have the same synchronous rotation before the impact. Following the impact, the rotation of the mantle is changed instantaneously. The field strength is estimated using the stationary equation (1), assuming a slow evolution of the differential rotation by comparison with the rapid growth of the dynamo and, hence, a quasistatic evolution of a saturated magnetic field. The depth to the Curie temperature of metallic iron, 1,040 K, is estimated by the conductive cooling of a semi-infinite half-space with initial and surface temperatures of 1,400 and 210 K, respectively.

31. Lavorel, G. & Le Bars, M. Experimental study of the interaction between convective and elliptical instabilities. *Phys. Fluids* **22**, 114101 (2010).
32. Cébron, D., Maubert, P. & Le Bars, M. Tidal instability in a rotating and differentially heated ellipsoidal shell. *Geophys. J. Int.* **182**, 1311–1318 (2010).
33. Greenspan, H. P. *The Theory of Rotating Fluids* (Cambridge Univ. Press, 1968).
34. Cébron, D., Le Bars, M. & Meunier, P. Tilt-over mode in a precessing triaxial ellipsoid. *Phys. Fluids* **22**, 116601 (2010).
35. Peale, S. J. in *Planetary Satellites* (ed. Burns, J. A.) 87–112 (Univ. Arizona Press, 1977).
36. Lacaze, L., Le Gal, P. & Le Dizès, S. Elliptical instability in a rotating spheroid. *J. Fluid Mech.* **505**, 1–22 (2004).
37. Correia, A. C. M. & Laskar, J. Mercury’s capture into the 3/2 spin–orbit resonance including the effect of core–mantle friction. *Icarus* **201**, 1–11 (2009).

Stepwise evolution of stable sociality in primates

Susanne Shultz¹, Christopher Opie¹ & Quentin D. Atkinson^{1,2}

Although much attention has been focused on explaining and describing the diversity of social grouping patterns among primates^{1–3}, less effort has been devoted to understanding the evolutionary history of social living⁴. This is partly because social behaviours do not fossilize, making it difficult to infer changes over evolutionary time. However, primate social behaviour shows strong evidence for phylogenetic inertia, permitting the use of Bayesian comparative methods to infer changes in social behaviour through time, thereby allowing us to evaluate alternative models of social evolution. Here we present a model of primate social evolution, whereby sociality progresses from solitary foraging individuals directly to large multi-male/multi-female aggregations (approximately 52 million years (Myr) ago), with pair-living (approximately 16 Myr ago) or single-male harem systems (approximately 16 Myr ago) derivative from this second stage. This model fits the data significantly better than the two widely accepted alternatives (an unstructured model implied by the socioecological hypothesis or a model that allows linear stepwise changes in social complexity through time). We also find strong support for the co-evolution of social living with a change from nocturnal to diurnal activity patterns, but not with sex-biased dispersal. This supports suggestions that social living may arise because of increased predation risk associated with diurnal activity. Sociality based on loose aggregation is followed by a second shift to stable or bonded groups. This structuring facilitates the evolution of cooperative behaviours⁵ and may provide the scaffold for other distinctive anthropoid traits including coalition formation, cooperative resource defence and large brains.

Anthropoids differ from other social vertebrates in the prevalence of stable groups and bonded relationships between individuals⁶. Explaining how primate social systems evolved is central to understanding the evolution of our closest relatives and the emergence of early human social behaviour⁷. Conventional explanations have appealed more to adaptive reasoning than phylogenetic history to account for patterns of sociality⁴. Adaptive arguments often invoke the socioecological model⁸, which predicts that individuals readily alter patterns of aggregation in response to ecological conditions^{3,9}. This focus has resulted in less emphasis on the historical processes and phylogenetic constraints that have informed other areas of evolutionary biology¹⁰.

However, behaviour, like morphology, physiology and life history, is heritable¹¹ and shaped by historical processes. Primate social behaviour is no exception; Old World primates, particularly cercopithecines, have highly inflexible social structures, and social traits cluster according to taxonomic grouping across the order⁴. Strong historical constraints make it crucial to incorporate phylogeny when testing adaptive explanations, but also create the possibility of explicitly modelling the evolutionary pathways leading to extant primate grouping patterns.

To evaluate the evolution of stable sociality in primates, we mapped the composition of foraging groups (solitary, family groups, harems or multi-male; see Supplementary Information for further discussion of alternative classification schemes) for 217 species onto a primate consensus tree (Fig. 1 and Supplementary Information) derived from genetic data¹². We then evaluated the strength of phylogenetic inertia

in the data (historical non-independence) using Pagel's lambda (λ)¹³. A λ value of 0 implies evolution independent of the phylogenetic tree, whereas a value of 1 indicates that the probability of shared inheritance between species is proportional to their relatedness. Social grouping patterns showed a strong phylogenetic signal ($\lambda_{\text{max}} = 0.983$, maximum likelihood ($LL_{\text{max}} = -150.038$) (significantly different from a λ value of 0 ($LL_0 = -332.63$, $P < 0.001$), but not significantly different from a λ value of 1 ($LL_1 = -141.12$, $P = 0.189$)). Flexible social structure is characteristic of only two groups, the Callitrichidae and Lemuridae (Fig. 1).

This strong phylogenetic signal allows a reconstruction of the evolutionary pathways leading to extant primate grouping patterns. Theoretical models suggest two possibilities. First, the socioecological model posits that grouping patterns are driven by individual responses to resource availability^{3,9}. Under this 'unstructured' model, if grouping patterns are facultative, transitions between all possible social states (and polymorphic states within species) should be equally likely. Second, primate social complexity has been proposed to increase in a stepwise fashion from solitary individuals, through small groups to large, socially complex groups^{14–16}. From this 'increasing complexity' model we would predict that pair-living was the earliest form of social group, followed by more complex grouping patterns. Support for such a model of social evolution through pair-bonds has been found in birds^{17,18} and insects¹⁹.

We used a Bayesian framework, implemented in BayesTraits²⁰, to evaluate four alternative models of social evolution (Fig. 2), including the two described above, on a posterior distribution of primate trees. The simplest model estimates a single rate of transition between all social states, representing an unstructured 'null' model of social change in which all state changes occur at the same underlying rate. We contrast this with a second, parameter-rich model in which rates are allowed to vary across all transitions. This model implies that some transitions are more likely than others, for example the rate from solitary to pair-living may be different from the rate from pair-living to solitary or to some other state—but does not make assumptions about what this structure will be. The third model simulates increasing complexity by restricting possible transitions to stepwise changes up and down a chain linking solitary to pair-living, to small harem groups and finally to large multi-male/multi-female groups. The fourth model is derived from the data and identifies likely transitions using the reversible-jump procedure in BayesTraits, which searches the posterior distribution of possible models by linking (setting to equal) or removing (setting to zero) transition rate parameters.

The model with the highest posterior support in the reversible-jump analysis (Supplementary Table 1) suggests that social evolution proceeds from solitary to multi-male/multi-female groups and then either to pair-living or harems. Back transitions occur from harems to multi-male groups, whereas transitions between pair-living and harems do not occur. Transitions from solitary to social are not reversed; such that once a lineage becomes social it remains so. We used Bayes Factors^{20–22} to test whether there is sufficient signal in the primate sociality data to support decisively any of the four alternative models. Table 1 shows that the reversible-jump-derived model is not only the

¹Institute of Cognitive and Evolutionary Anthropology, 64 Banbury Road, University of Oxford, Oxford OX2 6PN, UK. ²Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand.

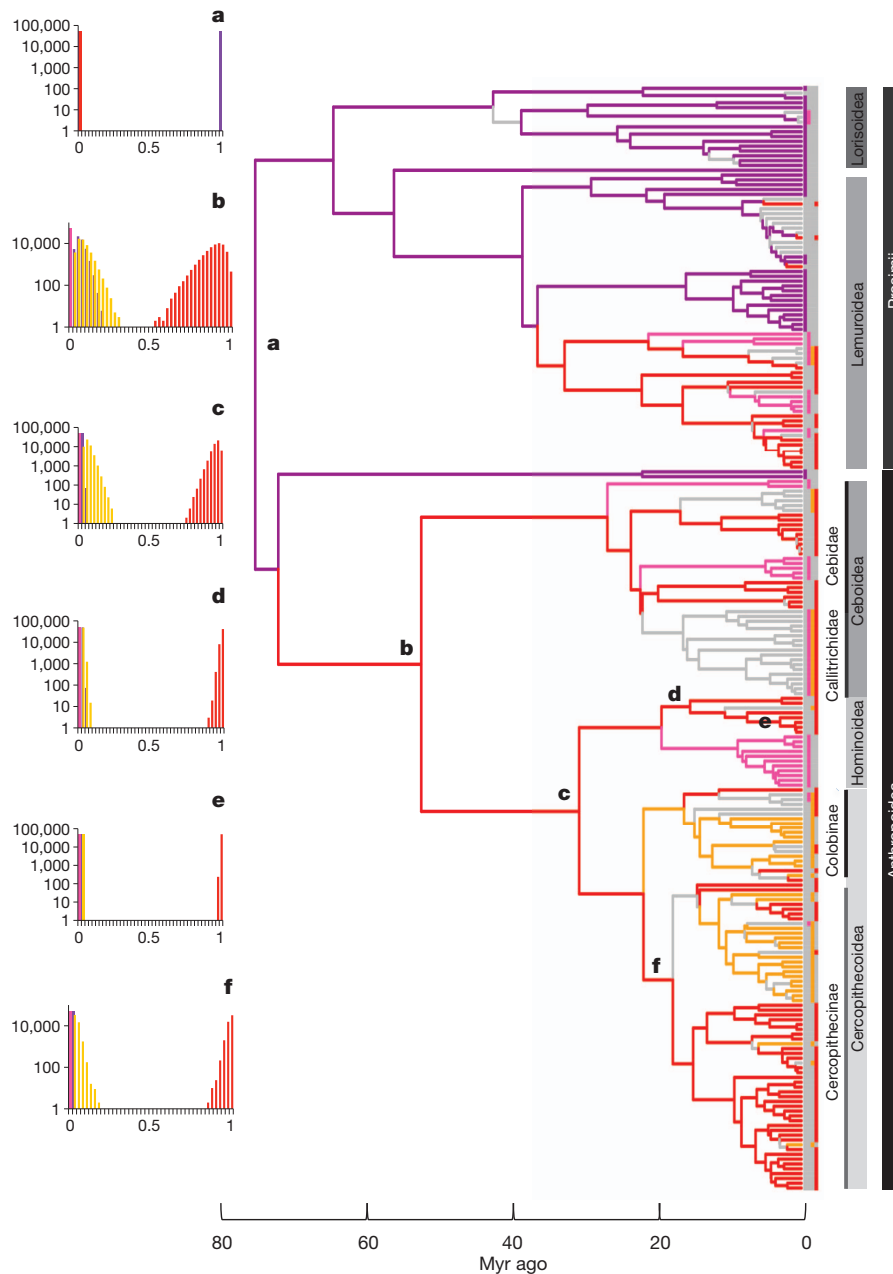


Figure 1 | Primate phylogeny showing ancestral state reconstructions for sociality under the reversible-jump Markov chain Monte Carlo-derived model of evolution. The tree topology is the maximum clade credibility tree from the 10kTrees Project¹² posterior distribution with branch lengths drawn proportional to time. Branches and tips are coloured for solitary (purple), unimale (orange), multi-male (red), pair-living (pink) where the combined

best fit to the data, but is also decisively better at explaining the data than the equal rates, the fully parameterized or the increasing complexity models.

We used the reversible-jump-derived model of social evolution to reconstruct the evolutionary history of social organization across the primate tree (Fig. 1). Ancestral node reconstructions reveal that the transition from solitary foraging at the primate root (74 Myr ago) to social aggregations was established at the anthropoid root (52 Myr ago) and the root of the Indriidae and Lemuridae (32 Myr ago) in prosimians. Other forms of social grouping evolved later in primates; harems appeared at the root of the Colobinae (16 Myr ago), followed soon after in the Cercopithecini (14 Myr ago). Pair-living arose at the root of the Callitrichidae (16 Myr ago), Hylobatidae (8.6 Myr ago), *Avahi* (6.4 Myr ago), hapalemurs (6.3 Myr ago), *Aotus* (4.8 Myr ago)

probability of the state and the branch is greater than or equal to 0.7. Where the combined probability is less than 0.7, the branch is grey. Histograms represent the posterior probability distribution of each social state at the nodes indicated (a, primate root; b, anthropoid root; c, catarrhine root; d, great ape root; e, *Pan-Homo* split; f, Old World monkey root).

and *Callicebus* (4.5 Myr ago). Thus, the fundamental shift to sociality occurred with the appearance of aggregations, followed later by derived grouping structures, including pair-living.

We next examined two possible catalysts of primate social evolution. First, the switch to social living is presumed to occur under increased predation pressure¹ coinciding with the shift from nocturnal to diurnal activity. We used a test of co-evolution in BayesTraits²⁰ to assess whether changes in activity patterns predict the major transition to social living. There was decisive support²² for the dependent model (that is, co-evolution between activity and sociality, Fig. 3a) over the independent model (mean $LL_{D(\text{dependent})} = -33.03 \pm 0.08$ s.e.m. compared with $LL_{I(\text{independent})} = -41.71 \pm 0.04$ s.e.m.; Bayes Factor 3.39; Supplementary Table 2), supporting the proposed link between the evolution of activity patterns and social living. Additionally, both

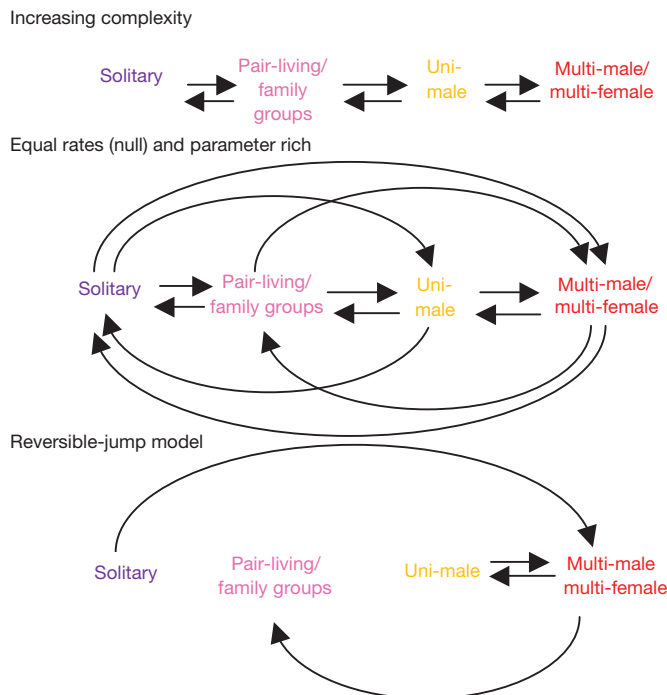


Figure 2 | Alternative evolutionary models of primate social evolution. Arrows represent allowable transitions between modes of social living under each model. Under the complexity and parameter-rich model, transition rates represented by each arrow can vary. Under the equal rates model, all rates are fixed to a single optimized rate parameter. The reversible-jump-derived model is a significantly better fit to the data than the alternative models.

intermediate states (social/nocturnal and solitary/diurnal) are unstable as the transition rate from these states to social/diurnal is an order of magnitude higher than any other transition. This suggests that the switch from a solitary, nocturnal lifestyle to diurnal social living represents a major shift in the primate adaptive landscape. Group living has long been argued to provide anti-predator benefits¹, and the shift to diurnal social living in primates would have opened up a vast new adaptive space in a highly visual world²³.

The second possible catalyst is the switch to sex-biased dispersal, whereby one sex (typically males) disperses further from the natal range than the other. This is assumed to be an ancestral or default mammalian characteristic²⁴. Changes in dispersal behaviour may be important in the evolution of sociality because in its extreme form, philopatry, one sex foregoes dispersal and remains in the natal range resulting in kin structured groups. A switch to sex-biased dispersal could therefore facilitate kin selection and the emergence of cooperative social groups^{25–27}. The extension of the mother–daughter bond to groups of related females also has been proposed as the fundamental relationship underpinning mammalian sociality²⁸. We used Discrete to evaluate whether sex-biased dispersal precedes the shift to sociality in primates. Although we find support for co-evolution between social grouping and dispersal patterns (mean $LL_D = -73.27 \pm 0.03$ s.e.m. versus $LL_I = -74.66 \pm 0.05$ s.e.m.; Bayes Factor 1.21; Fig. 3b), the association is much weaker than between sociality and activity patterns and independent models are sampled above chance (Supplementary Table 3).

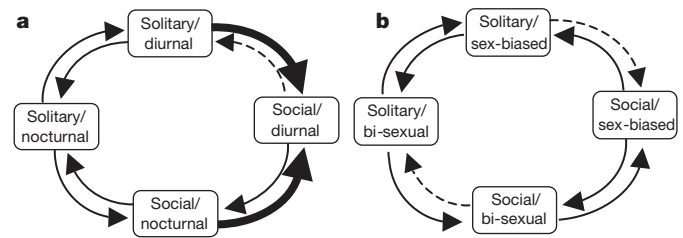


Figure 3 | Estimated transition rates for co-evolution of social living. Estimated transition rates with (a) activity and (b) dispersal patterns. Thin lines, an estimated median transition rate >0 but <0.01 ; heavy lines, a rate >0.01 ; dashed lines, a median estimated zero transition rate. Full estimated rates are reported in Supplementary Information.

Additionally, contrary to the assumption of sex-biased dispersal being a primate (and mammalian) default²⁴, the ancestral state for primates is bi-sexual dispersal (posterior probability of 0.93) and the estimated transition rates indicate that sex-biased natal dispersal follows the shift to sociality rather than precedes it (Fig. 3b).

Dispersal changes, therefore, do not trigger social living, but as they follow the emergence of social living they could be associated with a secondary transition to stable groups. A similar suggestion was put forward in a controversial model for the evolution of cooperative sociality in eusocial insects⁵. The model argues that aggregating individuals first create population structure. Stable groups then emerge secondarily through increased persistence resulting from silenced dispersal in at least one sex. To test whether this model explains the evolution of stable primate groups, we classified species as solitary, unstable social or stable social (the later defined as species with natal philopatry coupled with no/limited secondary dispersal or those with stable, long-term pair bonds). We then used the reversible jump procedure to identify the most likely model for the evolution of group stability; our model suggests that solitary living is the ancestral state, followed by unstable groups, and with a final transition to stable social groups (marginal $LL = -65.2 \pm 0.019$, Supplementary Table 4). This model is a better fit to the data than either an equal rates ($LL = -71.59 \pm 0.021$, Bayes Factor = 2.77) or a parameter-rich model, where transitions are allowed between all states ($LL = -69.08 \pm 0.052$; Bayes Factor = 1.66; Fig. 4). It thus appears that although the evolution of social groups does not occur through increasing complexity as defined by group size, there is strong support for a model of stepwise transitions leading from solitary living to unstable social aggregations, followed by a second step to stable groups based on either kinship or reproductive ties. Although transitions to social grouping are not uncommon in vertebrates, this secondary transition to stable grouping is, and may hold the key to the evolution of cooperative sociality characteristic of anthropoid primates, particularly humans.

Our analyses demonstrate a model of primate social evolution, which highlights the initial switch from solitary foraging to multi-male/multi-female aggregations. Although we cannot directly test adaptive explanations, our findings show this switch co-evolved with a change from a nocturnal to a diurnal lifestyle, supporting the role of predation in driving social evolution. Although group size has often been used as a proxy of social complexity in primates, relationship or group stability represents a more important indication of social

Table 1 | Comparison of alternative model performance

Model	Rank	Parameters	Mean likelihood	\log_{10} [Bayes Factor]
Reversible-jump Markov chain Monte Carlo-derived model	1	4	−64.84	–
Parameter-rich (unconstrained) model	2	12	−72.13	5.03
Equal rate ‘null’ model	3	1	−76.82	5.24
Increasing complexity	4	6	−77.55	6.5

Table shows number of model parameters, model rank, likelihood and \log_{10} [Bayes Factors] (see Supplementary Information). The Bayes Factor indicates relative support for the reversible-jump-derived model over alternatives (0–0.5 minimal; 0.5–1.0 substantial; 1.0–2.0 strong; >2.0 decisive)²⁹.

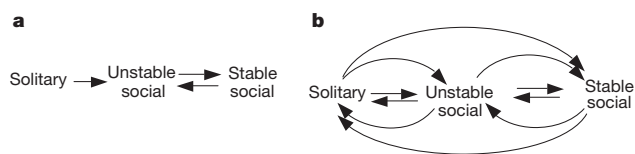


Figure 4 | Alternative evolutionary models for the evolution of stable grouping patterns. **a**, The model with the highest posterior support for the evolution of stable or bonded social groups. The model implies that stable social groups evolve from sociality through unstable social groups. **b**, Alternative models allow transitions between all possible states, either all constrained to the same rate (equal rates model) or allowed to vary (parameter-rich model).

complexity⁶. Our models suggest that the initial switch to sociality involved loose or unstable multi-male/multi-female aggregations (as exemplified in diurnal lemurs) followed by secondary transitions to bonded social relationships between mothers and daughters²⁸ (philopatry) or reproductive adults⁶ (pair-living). This secondary transition may be a key step towards facilitating cooperative social behaviour. In non-primates, social structuring is most commonly characterized by aggregations, with bonding associated with pair-living (for example, birds, ungulates, carnivores), and kin-based groups limited to a few taxa such as elephants and cetaceans^{6,29}. Kin-based structuring parallels that seen in eusocial insects. Testing these evolutionary models in other phyla would reveal whether the pathways suggested for primate evolution are more widely characteristic of cooperative sociality.

METHODS SUMMARY

To account for uncertainty in the underlying phylogeny, model testing was undertaken across a Bayesian posterior distribution of 10,000 ultrametric primate trees derived from genetic data as part of version 2 of the 10kTrees Project¹². The maximum clade credibility tree we present was inferred from the complete 10kTrees sample using TreeAnnotator³⁰. Pagel's lambda was estimated using the Ape and Geiger (see Supplementary Information) packages in R.

BayesTraits²⁰ uses an Markov chain Monte Carlo method to derive posterior distributions of log-likelihoods, the rate parameters of models of evolution, and trait values at ancestral nodes on the phylogeny. Transition rates between all states were constrained to be equal for the unstructured model (producing a simple one-parameter model). Rates were allowed to vary freely to parameterize the flexible model. All rates in the increasing complexity model except the forward and backward transitions between solitary/pair-living, pair-living/uni-male and uni-male/multi-male groups were restricted to zero. Model transition rates were also determined using the reversible-jump procedure in BayesTraits. Reversible-jump models were ranked in order of their posterior probability to identify the top ranked model. Model performance was compared using a log₁₀[Bayes Factor]²¹. Co-evolution between behavioural traits was assessed using the Discrete package in BayesTraits. Social organization was classed as solitary (0) or social (1) (including pair-living); activity pattern as nocturnal (0) and diurnal (1); natal dispersal as bi-sexual (0) or sex-biased (1) (see Supplementary Materials for further details).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 May; accepted 30 September 2011.

1. van Schaik, C. P. Why are diurnal primates living in groups? *Behaviour* **87**, 120–144 (1983).
2. Terborgh, J. & Janson, C. H. The socioecology of primate groups. *Annu. Rev. Ecol. Syst.* **17**, 111–135 (1986).
3. Sterck, H. M., Watts, D. P. & van Schaik, C. P. The evolution of female social relationships in nonhuman primates. *Behav. Ecol. Sociobiol.* **41**, 291–309 (1997).

4. Di Fiore, A. & Rendall, D. Evolution of social organization: a reappraisal for primates by using phylogenetic methods. *Proc. Natl Acad. Sci. USA* **91**, 9941–9945 (1994).
5. Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
6. Shultz, S. & Dunbar, R. I. M. The evolution of the social brain: anthropoid primates contrast with other vertebrates. *Proc. R. Soc. Lond. B* **274**, 2429–2436 (2007).
7. Elton, S. Forty years on and still going strong: the use of hominin-cercopithecoid comparisons in palaeoanthropology. *J. R. Anthropol. Inst.* **12**, 19–38 (2006).
8. Jarman, P. J. The social organisation of antelope in relation to their ecology. *Behaviour* **48**, 215–268 (1974).
9. Kappeler, P. M. & van Schaik, C. P. Evolution of primate social systems. *Int. J. Primatol.* **23**, 707–740 (2002).
10. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598 (1979).
11. Stirling, D. G., Réale, D. & Roff, D. A. Selection, structure and the heritability of behaviour. *J. Evol. Biol.* **15**, 277–289 (2002).
12. Arnold, C., Matthews, L. J. & Nunn, C. L. The 10kTrees website: a new online resource for primate phylogeny. *Evol. Anthropol.* **19**, 114–118 (2010).
13. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
14. Dunbar, R. I. M. in *Primate Males: Causes and Consequences of Variation in Group Composition* (ed. Kappeler, P. M.) 259–268 (Cambridge Univ. Press, 2000).
15. Nunn, C. L. & van Schaik, C. P. in *Reconstructing Behavior in the Primate Fossil Record* (eds Plavcan, J. M., Kay, R. F., Jungers, W. L. & van Schaik, C. P.) 159–216 (Plenum, 2001).
16. Müller, A. E. & Thalmann, U. Origin and evolution of primate social organisation: a reconstruction. *Biol. Rev. Camb. Philos. Soc.* **75**, 405–435 (2000).
17. Cornwallis, C. K., West, S. A., Davis, K. E. & Griffin, A. S. Promiscuity and the evolutionary transition to complex societies. *Nature* **466**, 969–972 (2010).
18. Shultz, S. & Dunbar, R. I. M. Life-history, social bonding and adaptive peak shifts in avian brain evolution. *Biol. J. Linn. Soc.* **100**, 111–123 (2010).
19. Hughes, W. O. H., Oldroyd, B. P., Beekman, M. & Ratnieks, F. L. W. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**, 1213–1216 (2008).
20. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825 (2006).
21. Rambaut, A. & Drummond, A. J. Tracer v1.4. *Tracer* (<http://beast.bio.ed.ac.uk/Tracer>) (2007).
22. Kass, R. E. & Raftery, A. E. Bayes Factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
23. Barton, R. A. Visual specialization and brain evolution in primates. *Proc. R. Soc. Lond. B* **265**, 1933–1937 (1998).
24. Greenwood, P. J. Mating systems, philopatry and dispersal in birds and mammals. *Anim. Behav.* **28**, 1140–1162 (1980).
25. Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, 1998).
26. Hamilton, W. D. The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**, 17–52 (1964).
27. Gompper, M. E., Gittleman, J. L. & Wayne, R. K. Genetic relatedness, coalitions and social behaviour of white-nosed coatis, *Nasua narica*. *Anim. Behav.* **53**, 781–797 (1997).
28. Broad, K. D., Curley, J. P. & Keverne, E. B. Mother-infant bonding and the evolution of mammalian social relationships. *Phil. Trans. R. Soc. B* **361**, 2199–2214 (2006).
29. Shultz, S. & Dunbar, R. I. M. Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proc. Natl Acad. Sci. USA* **107**, 21582–21586 (2010).
30. Drummond, A. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.S. is supported by a Royal Society Dorothy Hodgkin Fellowship. We thank R.I.M. Dunbar for comments on the manuscript.

Author Contributions S.S. designed the study, compiled the data and executed analyses. C.O. executed analyses. Q.A. was involved in study design and advised on statistical analyses. All authors contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.S. (susanne.shultz@anthro.ox.ac.uk).

METHODS

Primate data. Primates were classified as solitary, pair-living and group-living⁹; group-living were further split into single and multi-male groups (Supplementary Fig. 1). Data were compiled mainly from secondary literature or review articles^{3,16,31–35} and several online sites (pin.primates.wisc.edu, www.theprimata.com). The species were coded as follows: solitary ($n = 40$), pair-living ($n = 53$), single-male, multi-female ($n = 67$) or multi-male, multi-female ($n = 121$). Dispersal was classified as male-biased ($n = 86$), female-biased ($n = 14$) or bi-sexual dispersal ($n = 105$). Recent papers have argued that dispersal is more flexible than classification schemes acknowledge³⁶. However, here we attempt to capture the characteristic dispersal behaviour for each species. Activity was classified as diurnal or nocturnal; cathemeral species were classed as polymorphic for activity. Species were also classified in multiple states when variation between or within populations was reported. One classification decision we faced was how to categorize species that spent most of their time foraging solitarily but were either known to have extended and stable social groups or had stable sleeping associations (for example, *Loris*, *Microcebus*, *Galago*, *Pongo*). Although no primate is truly solitary, these species are particularly problematical as a few well-documented studies suggest stable community structures in nocturnal species, yet they do not form stable foraging parties³⁷. The same discussion about whether orang-utans are social or solitary has longed plagued primatologists. For this reason, we used multiple classifications for these species: (1) solitary foragers, (2) polymorphic (solitary foraging plus social category), and (3) solitary foraging except for *Pongo*. This way we were able to evaluate the impact their classification had on model performance. The classification scheme that primarily relied on social grouping classification, with the exception of *Pongo*, had the highest mean likelihood (LL = -64.80), followed by the polymorphic classification scheme (LL = -66.14), and finally the scheme that classified *Pongo* as solitary (LL = -71.71). *Pongo* classification affects model fit as they would be the only example of an anthropoid primate to revert from social to solitary living. We evaluated the posterior probability of predicted rate classes (zero versus non-zero) for each transition across all three classification schemes (Supplementary Fig. 2). Finally, we classified stability based on reported adult dispersal or migration events for both males (typically secondary dispersal after joining new groups) and females (classified as post-partum dispersal). This classification is more subjective than the previous traits as the data are limited and often descriptive. We classified pair-living species as stable if group turnover events were typically associated with death or severe injury to one of the adults (rather than regular emigration by resident adults). For group-living species, we defined stability as at least one sex typically remaining in the group throughout adulthood (resulting in kin-based groups). The primary references that the classifications were based on are found in the Supplementary Table 6.

Tree. The primate phylogeny was based on a sample of 10,000 ultrametric trees from version 2 of the 10kTrees Project¹². This provides a posterior distribution of phylogenies using Bayesian inference from six mitochondrial (CYTB, COX1, COX2, 12S rRNA, 16S rRNA and a gene cluster) and three autosomal genes (MC1R, CCR5, SRY) for 230 primate species. The nodes of the consensus tree are dated using mean molecular branch lengths from the Bayesian analysis and six known fossil calibration points¹². The consensus tree is a maximum credibility tree and was inferred from the complete 10,000 tree sample using TreeAnnotator³⁰. As BayesTraits²⁰ (<http://www.evolution.rdg.ac.uk/SoftwareMain.html>) allows missing data, we included all species from the tree block rather than pruning the tree to fit the data.

Phylogenetic signal. Phylogenetic signal in data indicates that related species are more similar in a particular trait than would be expected by chance (that is, the trait of a daughter species is not independent of that of the parent). To quantify phylogenetic signal in our primate sociality data, we used the fitDiscrete function in the Geiger³⁸ package in R to calculate the maximum likelihood value of Pagel's lambda^{13,39} on the maximum credibility tree. A λ value of 1 is consistent with a model of evolution along the phylogeny (that is, a probability of shared inheritance proportional to relatedness), whereas a λ value of 0 suggests evolution independent of the phylogenetic tree⁴⁰. A likelihood ratio test was used to compare the fitted maximum likelihood value of λ with a model implying no phylogenetic signal ($\lambda = 0$) to a model of evolution along the tree ($\lambda = 1$). The likelihood ratio test follows a χ^2 distribution, with one degree of freedom. Polymorphisms were collapsed such that flexible species were assigned an additional flexible social category.

Model settings and performance. To identify the model best supported by the data for each analysis (social evolution, stability, social-activity and social-dispersal models, plus the social-stability data sets), we used the Discrete and Multistate option in BayesTraits²⁰. We began with the reversible-jump procedure, using a uniform hyper before seed exponential rate priors with mean and variance ranging between 0 and 2 (ref. 41). We initially explored using a uniform hyper-prior to seed exponential rate priors with mean and variance ranging between 0

and 2. Model performance was robust to choice of hyper prior. 'Rate dev' settings were set to achieve acceptance values within 20–40% (for most models this was 0.02, 0.05 or 0.1). To establish whether the models had converged, we evaluated the posterior distribution and trace of harmonic mean log-likelihoods; we assumed convergence when this distribution was approximately normal, the likelihood traces did not show large jumps across runs. Models visited by the Markov chain were ranked in order of their posterior probability (Supplementary Tables 1–4). The posterior sample of transition rates for the social evolution model is shown in Supplementary Fig. 3.

Each Markov chain Monte Carlo simulation was run five times for 30 million iterations sampled every 100, with the first 25 million iterations discarded as the burn-in period. Examination of the post-burn-in log-likelihood and rate parameters across the Markov chain plotted in Tracer²¹ indicated that runs had reached convergence by this time (25 million iterations) and effective sample sizes for the parameters of interest were all above 2,000. We report the posterior distribution for rate parameters, marginal log-likelihoods²¹ and states at ancestral nodes from the run with the median likelihood.

Model comparison: social evolution. We constructed four different models of social organization. First, all rates were set equal, simulating equal likelihood for all transitions. Second, rates were allowed to vary freely without constraint to produce a 'flexible' model. Third, we ran a 'complexity' model where transitions were restricted so that movements were only allowed between solitary and pair-living, pair-living and uni-male harems, and uni-male harems and multi-male social organization. Finally, the model structure with the highest posterior support from the reversible-jump analysis was run, allowing transitions from solitary to multi-male and from multi-male to pair-living and to uni-male and back. All other rates were set to zero. Final models were run using uniform rate priors (0–0.3) across a range informed by either the reversible-jump analyses for the data driven models or maximum likelihood analyses for theoretical models. Examination of posterior distributions indicated that the rates were well within the prior bounds. Stability of the models was checked by evaluating variance in the mean log-likelihood values over five iterations of the final analyses.

To compare alternative models of social evolution, we calculated both the marginal likelihood and Bayes Factor (the ratio of the marginal likelihoods) using Tracer²¹. The Bayes Factor (BF) shows the weight of evidence to support one model over another, from 0 to 0.5 (minimal), to 0.5–1.0 (substantial), to 1.0–2.0 (strong), to greater than 2.0 (decisive)²².

Ancestral states. We used BayesTraits to infer the posterior probability of social behaviours at each ancestral node in the primate tree under the model with the highest posterior probability from the reversible-jump analysis. Although the results presented in Supplementary Fig. 1 are drawn on the maximum clade credibility tree, the analysis was performed across the posterior distribution of 10,000 primate trees. The ancestral state probabilities for each branch of the tree are the combined posterior probability of each state on that branch with the posterior probability that the branch itself exists.

Correlated evolution. The Discrete package in BayesTraits enables analysis of the co-evolution of two binary traits over a phylogeny. We ran two Discrete analyses to test the hypotheses that either dispersal or activity patterns determine social organization in primates by investigating the correlation and relative timing of changes in social organization with those in dispersal and activity. We ran the Discrete analysis with social organization as solitary (0) or social (1) (including pair-living), dispersal as bi-sexual (0) or either female or male (1) and activity as nocturnal (0) or diurnal (1) with cathemeral as (01). Model parameters and performance were established using the procedures described above. Exponential rate priors were seeded from a uniform hyper prior with mean and variance ranging between 0 and 2 (ref. 41). The posterior sample of reversible-jump Markov chain Monte Carlo models for social-activity analyses is shown in Supplementary Table 2, and for social-dispersal analyses in Supplementary Table 3. A Bayes Factor²¹ comparison was made between the independent and the dependent reversible-jump hyperprior model runs such that independent evolution could be rejected if there was support for the dependent model. In addition to a Bayes Factor comparison, we also investigated the number of visits to independent models in the dependent run to assess whether this was above chance²⁰. Mean and median transition rates for the two dependent analyses are reported in Supplementary Table 5.

31. Campbell, C. J., Fuentes, A., MacKinnon, K. C., Panger, M. & Bearder, S. K. *Primates in Perspective* (Oxford Univ. Press, 2007).
32. Dixon, A. *Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Human Beings* (Oxford Univ. Press, 1998).
33. Dunbar, R. I. M. *Primate Social Systems* (Chapman & Hall, 1988).
34. Smuts, B., Cheney, D., Seyfarth, R., Wrangham, R. & Struhsaker, T. *Primate Societies* (Univ. Chicago Press, 1987).
35. Isbell, L. A. & van Vuren, D. Differential costs of locational and social dispersal and their consequences for female group-living primates. *Behaviour* **133**, 1–36 (1996).

36. Jack, K. M. & Isbell, L. A. Dispersal in primates: advancing an individualized approach. *Behaviour* **146**, 429–436 (2009).
37. Schulke, O. & Ostner, J. Big times for dwarfs: social organization, sexual selection, and cooperation in the Cheirogaleidae. *Evol. Anthropol.* **14**, 170–185 (2005).
38. Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E. & Challenger, W. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131 (2008).
39. Pagel, M. D. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* **255**, 37–45 (1994).
40. Freckleton, R. P., Harvey, P. H. & Pagel, M. D. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
41. Pagel, M. D., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684 (2004).

Genome sequencing reveals insights into physiology and longevity of the naked mole rat

Eun Bae Kim^{1*}, Xiaodong Fang^{2*}, Alexey A. Fushan^{1*}, Zhiyong Huang^{2*}, Alexei V. Lobanov³, Lijuan Han², Stefano M. Marino³, Xiaoqing Sun², Anton A. Turanov³, Pengcheng Yang², Sun Hee Yim³, Xiang Zhao², Marina V. Kasaikina³, Nina Stoletski³, Chunfang Peng², Paz Polak³, Zhiqiang Xiong², Adam Kiezun³, Yabing Zhu², Yuanxin Chen², Gregory V. Kryukov^{3,4}, Qiang Zhang², Leonid Peshkin⁵, Lan Yang², Roderick T. Bronson⁶, Rochelle Buffenstein⁷, Bo Wang², Changlei Han², Qiye Li², Li Chen², Wei Zhao², Shamil R. Sunyaev^{3,4}, Thomas J. Park⁸, Guojie Zhang², Jun Wang^{2,9,10} & Vadim N. Gladyshev^{1,3,4}

The naked mole rat (*Heterocephalus glaber*) is a strictly subterranean, extraordinarily long-lived eusocial mammal¹. Although it is the size of a mouse, its maximum lifespan exceeds 30 years, making this animal the longest-living rodent. Naked mole rats show negligible senescence, no age-related increase in mortality, and high fecundity until death². In addition to delayed ageing, they are resistant to both spontaneous cancer and experimentally induced tumorigenesis^{3,4}. Naked mole rats pose a challenge to the theories that link ageing, cancer and redox homeostasis. Although characterized by significant oxidative stress⁵, the naked mole rat proteome does not show age-related susceptibility to oxidative damage or increased ubiquitination⁶. Naked mole rats naturally reside in large colonies with a single breeding female, the 'queen', who suppresses the sexual maturity of her subordinates⁷. They also live in full darkness, at low oxygen and high carbon dioxide concentrations⁸, and are unable to sustain thermogenesis⁹ nor feel certain types of pain^{10,11}. Here we report the sequencing and analysis of the naked mole rat genome, which reveals unique genome features and molecular adaptations consistent with cancer resistance, poikilothermy, hairlessness and insensitivity to low oxygen, and altered visual function, circadian rhythms and taste sensing. This information provides insights into the naked mole rat's exceptional longevity and ability to live in hostile conditions, in the dark and at low oxygen. The extreme traits of the naked mole rat, together with the reported genome and transcriptome information, offer opportunities for understanding ageing and advancing other areas of biological and biomedical research.

We applied a whole-genome shotgun strategy to sequence the genome of an individual male naked mole rat (NMR) (Table 1 and Supplementary Tables 1–3). The sequencing depth of 98.6% of the

genome assembly was more than 20-fold (Supplementary Figs 1–4). The mitochondrial genome was also assembled. Approximately 25% of the NMR genome was represented by transposon-derived repeats, which is lower than in other mammals (40% in human, 37% in mouse, and 35% in rat genomes) (Supplementary Tables 4 and 5, Supplementary Figs 5–7). The predicted NMR gene set included 22,561 genes (Table 1 and Supplementary Table 6), which is comparable to other mammals (22,389 in human, 23,317 in mouse, and 22,841 in rat). Of these, 21,394 (94.8%) genes were transcribed (based on the RNA-seq data for seven organs). More than 98% of NMR genes could be functionally annotated using homology approaches (Supplementary Table 7), and the quality of predicted genes was comparable to that of well-annotated mammalian genomes (Supplementary Tables 6 and 8 and Supplementary Fig. 8).

Most of the NMR genome (93%) showed synteny to human, mouse or rat genomes (Supplementary Table 9), and pairwise comparisons suggested a relatively low rate of NMR genome rearrangements after the split from the murid common ancestor. We defined common synteny blocks in human, mouse, rat and NMR genomes and identified segmental duplications and lineage-specific insertions and deletions (Supplementary Tables 10 and 11 and Supplementary Fig. 9). By analysing single-copy orthologous groups, we constructed a phylogenetic tree involving the NMR and other mammals (Fig. 1). As expected, the NMR placed within rodents and its ancestor split from the ancestor of rats and mice approximately 73 million years ago, whereas the ancestor of NMR, mouse and rat split from rabbit approximately 86 million years ago. Thus, in spite of some exceptional traits, the overall properties of the NMR genome appeared to be similar to those of other mammals.

Lineage-specific gene family expansions may be associated with the emergence of specific functions and physiology. Compared to other mammals, the NMR showed a moderate number of gene families under expansion and contraction (Fig. 1b), including 96 NMR lineage-specific gene families (Fig. 2). Analysis of syntenic regions identified 750 gained and 320 lost NMR genes (Supplementary Tables 12–14). At least 75.5% of genes gained showed evidence of transcription, and the lost genes were enriched for ribosome and nucleoside biosynthesis functions (Supplementary Table 15). We also identified 244 pseudogenes, containing 183 frameshift and 119 premature termination events (Supplementary Tables 16 and 17). Functional categories enriched for pseudogenes included olfactory receptor activity (GO:0004984, $P < 0.001$, Fisher's exact test, 36 genes), visual perception (GO:0007601, $P = 0.015$, *CRB1*, *CRYBB3*, *GNAT2*, *GRK7*, *GUCA1B* and *PDE6H*), spermatogenesis

Table 1 | Global statistics of the NMR genome

Sequencing	Insert size (bp)	Total data (Gb)	Sequence coverage (fold)
Paired end libraries	170–800	126.52	47
	2–20 × 10 ³	120.66	45
	Total	247.18	92
Assembly	N50 (kb)	Longest (kb)	Size (Gb)
Contigs	19.3	179	2.45
Scaffolds	1,585	7,787	2.66
Annotation	Number	Total length (Mb)	Percentage of the genome
Repeats	3,090,116	666.7	25
Genes	22,561	722.3	27.1
CDS	181,641	32.5	1.2

¹Department of Bioinspired Science, Ewha Womans University, Seoul, 120-750, Korea. ²BGI-Shenzhen, Shenzhen, 518083, China. ³Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁵Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁶Rodent Histopathology Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Department of Physiology and The Sam and Ann Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, Texas 78245, USA. ⁸Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA. ⁹Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark. ¹⁰Department of Biology, University of Copenhagen, Copenhagen, DK-2200 Copenhagen N, Denmark.

*These authors contributed equally to this work.

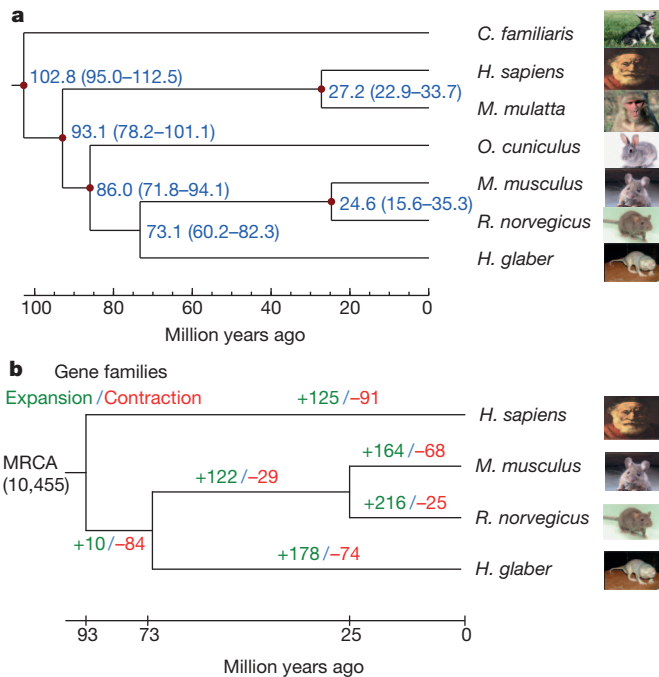


Figure 1 | Relationship of the NMR to other mammals. **a**, Estimation of the time of divergence (with error range shown in parentheses) of the NMR and six other mammals based on orthology relationship. Distances are shown in millions of years. **b**, Expansion and contraction in gene families. Numbers designate the number of gene families that have expanded (green) and contracted (red) since the split from the common ancestor. The most recent common ancestor (MCRA) has 10,455 gene families.

(GO:0007283, $P = 0.044$, *ADAM29*, *ADC*, *CCIN*, *CCT6B*, *DEDD*, *OAZ3* and *SHBG*), and possibly RING domain (SM00184, $P = 0.142$, *CNOT4*, *KCNRG*, *RNF5*, *TRIM17*, *TRIM11* and *ZSWIM2*). The enrichment in the visual perception category appears to underlie the evolution of poor vision in the NMR, whereas many RING-domain-containing proteins act as ubiquitin ligases¹². The levels of ubiquitinated proteins in NMRs are lower than in mice and, unlike those in mice, do not change significantly with age⁶.

Identification of genes that have undergone positive selection in the NMR lineage can provide useful pointers to the evolution of its unique

traits. 45 genes (0.4%) were identified as positively selected in the NMR lineage at the false discovery rate of 0.01 and 141 genes (1.2%) at the false discovery rate of 0.05 (Supplementary Table 18). 12 out of the 45 genes (corresponding to the false discovery rate of 0.01) passed a strict manual inspection for alignment quality. In comparison, 0.7% of genes were predicted to be positively selected in the human lineage from high-quality alignments and using Rom correction for multiple testing¹³. Interestingly, our set included *TEP1*, encoding a telomerase component, and *TERF1*, a telomeric repeat binding factor identified at the false discovery rate of 0.05 (Supplementary Fig. 10). The *TERF1* gene product is one of six proteins contributing to the shelterin complex, which shapes and protects telomeres¹⁴ and has been proposed to regulate telomere length¹⁵.

To gain further insights into biological processes that underlie the exceptional traits of the NMR, we identified 39 NMR proteins containing 45 amino acid residues unique among orthologues present in 36 vertebrate genomes (Supplementary Table 19). This gene set included cyclin E1 (*CCNE1*), uncoupling protein 1 (*UCP1*) and γ -crystallin (*CRYGS*), which are associated with the G1/S transition during the cell cycle, thermogenesis and visual function, respectively. Other noteworthy genes were *APEX1*, a multifunctional DNA repair enzyme, *RFC1*, replication factor C, and *TOP2A*, a DNA topoisomerase that controls the topologic states of DNA during transcription. This set also contained eight genes designated as cancer-related¹⁶. Finally, *TOP2A*, along with *TEP1* and *TERF1* from the set of positively selected genes, are part of a five-protein complex of alternate lengthening of telomere pathway¹⁷. Overall, these analyses point to altered telomerase function in the NMR, which may be related to its evolution of extended lifespan and cancer resistance.

We also identified 1.87 million heterozygous single-nucleotide polymorphisms (SNPs). This results in an estimated nucleotide diversity (mean per nucleotide heterozygosity) of 7×10^{-4} , which is much lower than in mouse and rat populations and is comparable to the nucleotide diversity observed in humans. Transition nucleotide changes were observed twice as often as transversions, indicating that variant calls reproduce the expected properties of natural variation in other mammals. This low level of nucleotide diversity may reflect a low effective size of NMR population, but may also be due to a high level of inbreeding, a reduced mutation rate or high efficiency of the repair systems. The variation of diversity along the genome was consistent with inbreeding in the NMR population. In protein-coding regions of the genome, our analysis identified 10,951 non-synonymous and 8,616 synonymous SNPs. Their ratio is much higher than in other studied organisms, including human, which appears to signal relaxation of purifying selection in the NMR, potentially as a consequence of reduced effective population size. Finally, we analysed the context dependency of NMR SNPs (Supplementary Fig. 11). Relative rates of nucleotide changes and nucleotide context dependencies were similar to those observed in human polymorphism, with the exception of a relative reduction of SNPs due to CpG mutations. This was caused by a combination of the relatively low CpG density in the NMR genome and a higher fraction of CpG dinucleotides within CpG islands compared to the human genome. CpG density was only 0.19 of that expected on the basis of the GC content, which is lower than in human, dog and panda genomes, but is similar to the mouse genome. However, in comparison to mouse, a higher fraction of CpG dinucleotides was concentrated in CpG islands. CpG dinucleotides within CpG islands contribute less to genetic variation because of their lower methylation rate and possibly also due to selection.

Long lifespan is a key feature of the NMR. To study ageing and longevity, we obtained RNA-seq data for brain, liver and kidney of newborn, young adult (4-year-old) and old adult (20-year-old) NMRs (Supplementary Table 20). In contrast to other mammals, few genes showed differential expression between 4- and 20-year-old NMRs, especially in the brain (Supplementary Tables 21–23). A recent study identified 33 underexpressed and 21 overexpressed genes in the human brain during ageing¹⁸. Of these, 32 genes did not show

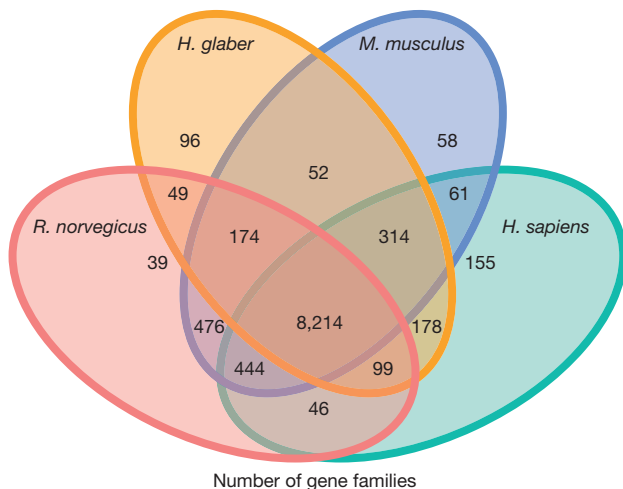


Figure 2 | Common and unique NMR gene families. This Venn diagram shows unique and overlapping gene families in the NMR (*H. glaber*), rat (*R. norvegicus*), mouse (*M. musculus*) and human (*H. sapiens*).

consistent expression changes with ageing in NMRs, including 30 genes that had stable expression and two genes that changed in the opposite direction compared to human brain (Supplementary Table 21). For example, *CYP46A1* and *SMAD3* were downregulated in the human brain, but showed elevated expression in the NMR brain. The product of the *CYP46A1* gene is a mediator of cholesterol homeostasis that influences the tendency of A β to aggregate. The product of *SMAD3* is a modulator of TGF- β signalling, playing a role in cancer development by slowing down the rate of cell proliferation. Elevated expression of *SMAD3* in the NMR during ageing may help optimize the rate of cell death, protecting NMRs from cancer.

A previous meta-analysis of age-related gene expression in mice, rats and humans revealed 56 consistently overexpressed and 17 under-expressed genes¹⁹. However, many of these genes did not show the same expression changes, suggesting that different regulatory mechanisms may underlie NMR longevity (Supplementary Tables 22 and 23). For example, genes related to degradation of macromolecules, such as *GSTA1*, *DERL1* and *GNS*, were not upregulated with age in NMRs. We also found that genes encoding mitochondrial proteins (*NDUFB11*, *ATP5G3* and *UQCRCQ*) were not downregulated, consistent with stable maintenance of mitochondrial function during ageing. It is also of interest that *TERT* (telomerase reverse transcriptase) showed stable expression regardless of age (Supplementary Fig. 12). This finding is consistent with the role of the telomerase complex, highlighted by positive selection on *TEP1* and *TERF1*. Overall, transcriptome and sequence data revealed different (compared to humans, mice and rats) patterns of NMR genes, which may underlie longevity mechanisms in this animal.

Non-shivering thermogenesis is a major heat production process in mammals that mainly depends on the action of *UCP1*, one of the 39 vertebrate genes that changed uniquely in the NMR (Supplementary Table 19). *UCP1* featured changes in amino acids Gln 146, Arg 263, Trp 264 and Thr 303, with the latter two residues being subject to positive selection ($P < 0.05$, likelihood ratio test for the branch-site model, $n = 30$) and Arg 263 and Trp 264 located in the conserved nucleotide binding motif (Fig. 3a). With Arg–Trp instead of the rigid Gly–Pro in the key regulatory site, *UCP1* is expected to lose the tight regulation by purine nucleotides as inhibitors and fatty acids as activators (Fig. 3b and c). The same loop also features two positively

charged Lys residues followed by a negatively charged residue (also a unique combination), that should markedly affect the local electrostatic potential of *UCP1*. In addition, Gln 146 replaced a conserved His involved in proton transport, and the same mutation was shown to decrease proton conductance of *UCP1* fivefold²⁰. Thr 303 is located in the carboxy-terminal motif (RqTxDCxT) required for binding purine nucleotides²¹. Taken together, these observations indicate a tight association of *UCP1* function with the unique thermoregulation of the NMR²².

In mammals, switches between light and dark periods affect synthesis of the hormone melatonin, which modulates sleep and circadian rhythms. NMRs live in a naturally dark habitat and their pineal glands, where melatonin is synthesized, are atrophied²³, but we found that the genes involved in melatonin synthesis (*TPH1*, *TPH2*, *DDC*, *AANAT* and *ASMT*) are intact. Interestingly, the expression of genes involved in the final two steps of melatonin synthesis was very low (*AANAT*) or undetectable (*ASMT*) in the NMR brain regardless of age (Supplementary Table 24 and Supplementary Fig. 13). Moreover, two major mammalian melatonin receptors (*MTNRIA* and *MTNR1B*, encoding MT₁ and MT₂, respectively) were inactivated by mutations that introduce premature stop signals (Supplementary Fig. 14). Synteny analyses showed that these pseudogenes corresponded to mouse *MTNRIA* and *MTNR1B*. Although melatonin signalling appears to be disrupted in the NMR, its circadian rhythms were maintained in terms of locomotor activity and body temperature when exposed to periodic light/dark changes²⁴. Our finding is consistent with a previous report that MT₁/MT₂ knockout mice maintained essentially normal circadian rhythms²⁵. These mice also showed decreased insulin secretion²⁵. Likewise, our transcriptome analysis of the NMR revealed decreased expression of genes involved in insulin/IGF-1 signalling in the liver compared to mice (Supplementary Fig. 15).

To explain the extraordinary resistance of the NMR to cancer³, a two-tier protective mechanism involving contact inhibition mediated by p16^{Ink4a} and p27^{Kip1} was proposed⁴. The involvement of p16^{Ink4a} is unusual, since humans and mice show only contact inhibition mediated by p27^{Kip1}. We analysed the gene locus and the transcriptome reads corresponding to tumour suppressors p16^{Ink4a} and p19^{Arf}. As in mice, the p16^{Ink4a} transcript consists of three exons (Supplementary Fig. 16). However, sequence similarity in the last exon is low, and two early stop codons in the second exon were predicted to result in a shorter, 14-kDa protein (Supplementary Fig. 17). The four ankyrin repeats were, however, intact and Thr69, a residue important for CDK6 binding, was conserved, so the function of the protein may be partially preserved (Supplementary Fig. 18). The p19^{Arf} transcript consists of two exons, but four stop codons in the second exon should lead to a shorter, 10-kDa protein (Supplementary Figs 19–21).

The NMR is also unique in that its skin and cutaneous C-fibres lack the neuropeptide Substance P, making this animal insensitive to certain types of pain^{10,11}. Our analysis revealed the presence of intact *TAC1* encoding Substance P. However, the NMR had a deletion in the core promoter region highly conserved among mammals (Supplementary Fig. 22). Thus, this neurotransmitter appears to be functional but may be under unique regulation.

We further examined the molecular basis for poor visual function and small eyes in the NMR. Of the four vertebrate opsin genes (*RHO*, *OPN1LW*, *OPN1MW* and *OPN1SW*), two (*OPN1LW* and *OPN1MW*) were missing (Table 2); this distinguishes the NMR from other rodents with dichromatic colour vision, such as mice, rats and guinea pigs. However, the NMR has intact *RHO* (rhodopsin) and *OPN4* (melanopsin), supporting the presence of rod-dominated retinæ and the capacity to distinguish light/dark cues. Of about 200 genes associated with visual perception (GO:0007601) in humans and mice, almost 10% were inactivated or missing in the NMR (Table 2 and Supplementary Fig. 23). These mammalian genes participate in crystallin formation, phototransduction in the retina, retinal development, dark adaptation, night blindness and colour vision. For at least ten of these

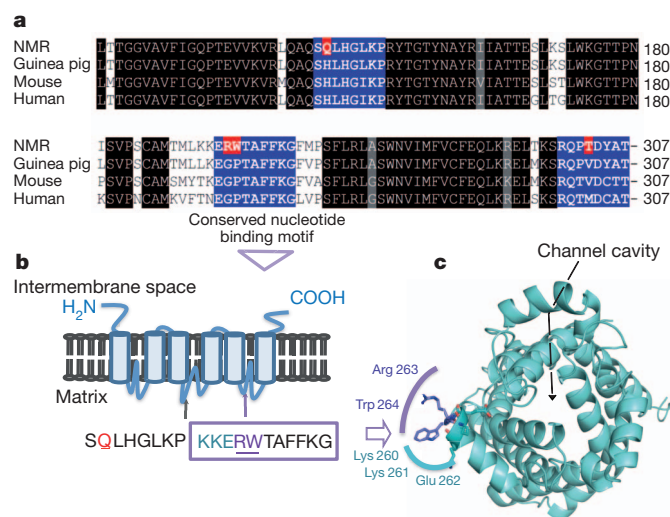


Figure 3 | Unique changes in UCP1 sequences and their roles in thermoregulation. **a**, Alignment of mammalian UCP1 sequences. Amino acids unique to the NMR are highlighted in red, and conserved motifs in blue. **b**, Topology of UCP1. Regions affected in the NMR are highlighted. **c**, Structural model of UCP1. Location of the channel and the nucleotide-binding loop with altered sequences in the NMR are shown.

Table 2 | Visual perception genes that are inactivated or are missing in the NMR genome

Gene	Inactivation event	Time of gene loss	ω_0 (average)	ω_1 (other)	ω_2 (NMR)	P-value
<i>RBP3</i>	F	NMR	0.121	0.085	0.232	1.187E-11
<i>ARR3</i>	F/S	NMR	0.420	0.260	0.912	6.263E-06
<i>PDE6C</i>	F	NMR	0.171	0.139	0.316	0.0001
<i>GUCA1B</i>	F	NMR	0.083	0.056	0.217	0.001
<i>GJA10</i>	F/S	NMR	0.308	0.248	0.524	0.002
<i>GUCY2E</i>	F	NMR	0.124	0.105	0.182	0.002
<i>CRYBA4</i>	S	NMR	0.055	0.036	0.123	0.001
<i>GNAT2</i>	F	NMR	0.055	0.039	0.108	0.017
<i>SLC24A1</i>	F	NMR	0.389	0.355	0.517	0.035
<i>CRYBB3</i>	S	NMR	0.071	0.048	0.122	0.037
<i>RP1L1</i>	S	NMR	0.448	0.424	0.513	0.186
<i>GRK7</i>	F	NMR	0.154	0.135	0.201	0.335
<i>PDE6H</i>	F	NMR	0.091	0.082	0.127	0.648
<i>EYS</i>	F/S	Ancestor	-	-	-	-
<i>GUCA1C</i>	F	Ancestor	-	-	-	-
<i>OPN1LW</i>	L	NMR	-	-	-	-
<i>OPN1MW</i>	L	NMR	-	-	-	-
<i>PRCD</i>	L	NMR	-	-	-	-
<i>RD3</i>	L	NMR	-	-	-	-

F, frameshift; S, premature stop codon; L, complete loss detected with synteny information. Ancestor/NMR indicates that the gene was lost in a rodent ancestor or the NMR, respectively. The rate ratio ω of non-synonymous to synonymous substitutions was calculated by using human, mouse, rat and NMR sequences. ω_0 is the average ratio in all branches, ω_1 is the average ratio in non-NMR branches, and ω_2 is the ratio in the NMR branch. A small P-value indicates that the two-ratio model fits the data better than the one-ratio model.

genes, we observed relaxation of the functional constrain on NMR sequences by estimating the ratio of non-synonymous to synonymous substitutions, which corroborated the dysfunction of these genes. Inactivation of *CRYBA4*, a microphthalmia-related gene, may be associated with the small-sized eyes, whereas inactivation of *CRYBA4* and *CRYBB3* and a NMR-specific mutation in *CRYGS* (Supplementary Table 19) may be associated with abnormal eye morphology²⁶. Thus, while some genes responsible for vision are preserved in the NMR, its poor visual function may be explained by deterioration of genes coding for various critical components of the visual system.

Further analysis revealed substantial divergence of the NMR nuclear receptor corepressor Hairless from other mammalian orthologues and the presence of amino acid replacements associated with the hairless phenotype, which is consistent with the lack of fur in NMRs (Supplementary Fig. 24). In addition, we found substantial sequence variation in the sweet taste receptor and lack of many bitter taste receptors common to other mammals (Supplementary Fig. 25 and 26). In particular, the NMR appears to lack the phenylthiocarbamide taste, a dominant genetic trait in humans, as well as several other common bitter tastes.

Air in NMR burrows is low in O₂ (~8%) and high in CO₂ (>10%) owing to many animals sharing a limited air supply and poor gas exchange through soil²⁷. To cope with the low O₂ conditions, the NMR has developed adaptive circulatory (altered haemoglobin oxygen affinity) and metabolic functions, reducing metabolic rate and slowing down development^{1,8,28,29}. To obtain insights into this adaptation, we examined gene expression changes in several tissues of NMR subjected to 8% O₂ for one week (Supplementary Tables 25–31 and Supplementary Fig. 27–30). Many changes associated with energy metabolism and redox control were observed. Sequence analysis of NMR hypoxia-induced factor 1 α (HIF1 α) revealed a T407I exchange unique among mammals and located in the VHL-binding domain. Under normal oxygen conditions, VHL mediates ubiquitin-dependent degradation of HIF1 α . In addition, NMR VHL harbours V166I exchange at a functionally important site. These amino acid changes are consistent with relaxation of ubiquitin-dependent degradation of HIF1 α , and, thus, with adaptation to low oxygen conditions.

To summarize, sequencing and analysis of the NMR genome revealed numerous insights into the biology of this remarkable animal. In addition, this genome and the associated data sets offer the research

communities working in ageing, cancer, eusociality and many other areas a rich resource that can be mined in numerous ways to uncover the molecular bases for the extraordinary traits of this most unusual mammal. In turn, this information provides unprecedented opportunities for addressing some of the most challenging questions in biology and medicine, such as mechanisms of ageing, the role of genetic makeup in regulating lifespan, adaptations to extreme environments, hypoxia tolerance, thermogenesis, resistance to cancer, circadian rhythms, sexual development and hormonal regulation.

METHODS SUMMARY

The NMR genome was sequenced on the Illumina HiSeq 2000 platform. The sequenced individual male NMR was from a captive breeding colony located at the University of Illinois, Chicago. The genome was assembled using SOAPdenovo. We obtained 2.5 Gb (gigabase pairs) contig sequences with N50 19.3 kb (kilobase pairs) and N90 4.7 kb, and 2.7 Gb scaffold sequences with N50 1.6 Mb (megabase pairs) and N90 0.3 Mb. (The N50 (or N90) contig size is the length of the smallest contig S in the sorted list of all contigs where the cumulative length from the largest contig to contig S is at least 50% (or 90%) of the total assembly length.) RNA-seq data (ageing and low O₂ experiments) were for animals from the same colony. See Supplementary Information for data analysis and additional details.

Received 14 June; accepted 5 September 2011.

Published online 12 October 2011.

- Edrey, Y. H., Park, T. J., Kang, H., Biney, A. & Buffenstein, R. Endocrine function and neurobiology of the longest-living rodent, the naked mole-rat. *Exp. Gerontol.* **46**, 116–123 (2011).
- Buffenstein, R. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *J. Comp. Physiol. B* **178**, 439–445 (2008).
- Liang, S., Mele, J., Wu, Y., Buffenstein, R. & Hornsby, P. J. Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (*Heterocephalus glaber*). *Aging Cell* **9**, 626–635 (2010).
- Seluanov, A. et al. Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proc. Natl Acad. Sci. USA* **106**, 19352–19357 (2009).
- Andziak, B. et al. High oxidative damage levels in the longest-living rodent, the naked mole-rat. *Aging Cell* **5**, 463–471 (2006).
- Pérez, V. I. et al. Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proc. Natl Acad. Sci. USA* **106**, 3059–3064 (2009).
- Jarvis, J. U. M. Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science* **212**, 571–573 (1981).
- Larson, J. & Park, T. J. Extreme hypoxia tolerance of naked mole-rat brain. *Neuroreport* **20**, 1634–1637 (2009).
- Buffenstein, R., Woodley, R., Thomadakis, C., Daly, T. J. & Gray, D. A. Cold-induced changes in thyroid function in a poikilothermic mammal, the naked mole-rat. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **280**, R149–R155 (2001).
- Park, T. J. et al. Selective inflammatory pain insensitivity in the African naked mole-rat (*Heterocephalus glaber*). *PLoS Biol.* **6**, e13 (2008).
- Smith, E. S., Blass, G. R., Lewin, G. R. & Park, T. J. Absence of histamine-induced itch in the African naked mole-rat and “rescue” by Substance P. *Mol. Pain* **6**, 29 (2010).
- Deshaies, R. J. & Joazeiro, C. A. P. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
- Schneider, A. et al. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol. Evol.* **1**, 114–118 (2009).
- de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
- van Steensel, B. & Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740–743 (1997).
- Higgins, M. E. et al. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.* **35**, D721–D726 (2007).
- Bhattacharyya, S. et al. Telomerase-associated protein 1, HSP90, and topoisomerase IIA associate directly with the BLM helicase in immortalized cells using ALT and modulate its helicase activity using telomeric DNA substrates. *J. Biol. Chem.* **284**, 14966–14977 (2009).
- Hong, M. G., Myers, A. J., Magnusson, P. K. E. & Prince, J. A. Transcriptome-wide assessment of human brain and lymphocyte senescence. *PLoS ONE* **3**, e3024 (2008).
- de Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of ageing. *Bioinformatics* **25**, 875–881 (2009).
- Bienengraeber, M., Echta, K. S. & Klingenberg, M. H⁺ transport by uncoupling protein (UCP-1) is dependent on a histidine pair, absent in UCP-2 and UCP-3. *Biochemistry* **37**, 3–8 (1998).
- Porter, R. K. Uncoupling protein 1: a short-circuit in the chemiosmotic process. *J. Bioenerg. Biomembr.* **40**, 457–461 (2008).

22. Woodley, R. & Buffenstein, R. Thermogenic changes with chronic cold exposure in the naked mole-rat (*Heterocephalus glaber*). *Comp. Biochem. Physiol. A* **133**, 827–834 (2002).
23. Quay, W. B. Pineal atrophy and other neuroendocrine and circumventricular features of the naked mole-rat, *Heterocephalus glaber* (Rüppell), a fossorial, equatorial rodent. *J. Neural Transm.* **52**, 107–115 (1981).
24. Riccio, A. P. & Goldman, B. D. Circadian rhythms of locomotor activity in naked mole-rats (*Heterocephalus glaber*). *Physiol. Behav.* **71**, 1–13 (2000).
25. Mühlbauer, E., Gross, E., Labucay, K., Wolgast, S. & Peschke, E. Loss of melatonin signalling and its impact on circadian rhythms in mouse organs regulating blood glucose. *Eur. J. Pharmacol.* **606**, 61–71 (2009).
26. Nikitina, N. V. et al. Postnatal development of the eye in the naked mole rat (*Heterocephalus glaber*). *Anat. Rec. A* **277A**, 317–337 (2004).
27. Bennett, N. C. & Faulkes, C. G. *African Mole-Rats: Ecology and Eusociality* (Cambridge University Press, 2000).
28. McNab, B. K. The influence of body size on the energetics and distribution of fossorial and burrowing mammals. *Ecology* **60**, 1010–1021 (1979).
29. Johansen, K., Lykkeboe, G., Weber, R. E. & Maloiy, G. M. Blood respiratory properties in the naked mole rat *Heterocephalus glaber*, a mammal of low body temperature. *Respir. Physiol.* **28**, 303–314 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. C. LaVinka for help in preparing NMR tissues and N. Estolaga for help in preparing the manuscript. We acknowledge financial support from the WCU Program (R31-2008-000-10010-0), the NIH (AG038004, AG021518 and CA080946), the Shenzhen Municipal Government (ZYC200903240077A), the

National Natural Science Foundation of China (30725008) and the National Science Foundation (0744979).

Author Contributions V.N.G. conceived the study. T.J.P. carried out animal work. A.A.T., M.V.K. and S.H.Y. prepared samples. X.F., Z.H., L.H., X.S., P.Y., X.Z., C.P., Z.X., Y.Z., Y.C., Q.Z., L.Y., B.W., C.H., Q.L., L.C., W.Z., G.Z. and J.W. performed genome sequencing and assembly. X.F., G.Z. and J.W. supervised genome sequencing and assembly. E.B.K., X.F., A.A.F., Z.H., A.V.L., S.M.M., L.P., G.Z. and V.N.G. performed genome and transcriptome analyses. Z.H., N.S., P.P., A.K. and S.R.S. carried out genetic analyses. G.V.K., R.T.B. and R.B. discussed the data. All authors contributed to data interpretation. V.N.G. wrote the paper with significant contributions from E.B.K., X.F., A.A.F., Z.H., S.R.S. and G.Z., and input from all authors.

Author Information The NMR whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number AFSB00000000. The version described in this paper is the first version, AFSB01000000. The mitochondrial sequence has been deposited at GenBank under the accession number JN242813. All short-read data have been deposited into the Short Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under the accession number SRA030468. Raw sequencing data of the transcriptome have been deposited in the Gene Expression Omnibus with the accession number GSE30337. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.Z. (zhanggj@genomics.org.cn), J.W. (wangj@genomics.org.cn) or V.N.G. (vgladyshev@rics.bwh.harvard.edu).

Active tactile exploration using a brain–machine–brain interface

Joseph E. O'Doherty^{1,2}, Mikhail A. Lebedev^{2,3}, Peter J. Ifft^{1,2}, Katie Z. Zhuang^{1,2}, Solaiman Shokur⁴, Hannes Bleuler⁴ & Miguel A. L. Nicolelis^{1,2,3,5,6}

Brain–machine interfaces^{1,2} use neuronal activity recorded from the brain to establish direct communication with external actuators, such as prosthetic arms. It is hoped that brain–machine interfaces can be used to restore the normal sensorimotor functions of the limbs, but so far they have lacked tactile sensation. Here we report the operation of a brain–machine–brain interface (BMBI) that both controls the exploratory reaching movements of an actuator and allows signalling of artificial tactile feedback through intracortical microstimulation (ICMS) of the primary somatosensory cortex. Monkeys performed an active exploration task in which an actuator (a computer cursor or a virtual-reality arm) was moved using a BMBI that derived motor commands from neuronal ensemble activity recorded in the primary motor cortex. ICMS feedback occurred whenever the actuator touched virtual objects. Temporal patterns of ICMS encoded the artificial tactile properties of each object. Neuronal recordings and ICMS epochs were temporally multiplexed to avoid interference. Two monkeys operated this BMBI to search for and distinguish one of three visually identical objects, using the virtual-reality arm to identify the unique artificial texture associated with each. These results suggest that clinical motor neuroprostheses might benefit from the addition of ICMS feedback to generate artificial somatic perceptions associated with mechanical, robotic or even virtual prostheses.

Brain–machine interfaces (BMIs) have evolved from 1-d.f. systems³ to many-d.f. robotic arms⁴ and muscle stimulators⁵ that perform complex limb movements, such as reaching^{6–8} and grasping⁹. However, somatosensory feedback, which is essential for dexterous control^{10–12}, remains underdeveloped in BMIs. With the exception of a few studies combining BMIs with tactile stimuli applied to the body¹³, existing systems rely almost exclusively on visual feedback. Prosthetic sensation has been studied in the context of sensory substitution¹⁴ and targeted reinnervation¹⁵; however, these approaches have limited application range and channel capacity. To provide a proof-of-concept method of equipping neuroprostheses with sensory capabilities, we implemented a BMBI that extracts movement commands from the motor areas of the brain while delivering ICMS feedback in somatosensory areas^{1,2,16} to evoke discriminable percepts^{17–20}. This idea received support from our pilot study¹⁶, in which a monkey responded to ICMS cues with the movements of a BMI-controlled cursor. However, the ICMS cue did not provide feedback of object–actuator interactions in this previous demonstration.

The BMBI developed here allowed active tactile exploration²¹ during BMI control (Fig. 1a). Two monkeys (M and N) received multielectrode implants in the primary motor cortex (M1) and the primary somatosensory cortex (S1) (Fig. 1b). They explored virtual objects using either a computer cursor or a virtual image (avatar) of an arm (Supplementary Fig. 1a, b). In 'hand control', the monkeys moved a joystick with their left hands to position the actuator. They searched through a set of virtual objects, selected one with a particular artificial

texture conveyed by ICMS, and held the actuator over that object to obtain reward (Fig. 1a and Supplementary Fig. 1c, d). During 'brain control', the joystick was disconnected and the actuator was controlled by the activity of right-hemisphere M1 neurons^{9,22,23}. The behavioural tasks varied in the number of objects on the screen, the artificial textures used and the actuator type (Fig. 2a), and were more difficult than previously reported BMI tasks because of the presence of multiple objects in the workspace, a prolonged object selection period and the necessity of interpreting ICMS feedback.

ICMS was delivered through two pairs of microwires to the hand representation area of S1 in monkey M (Fig. 1c) and through one pair

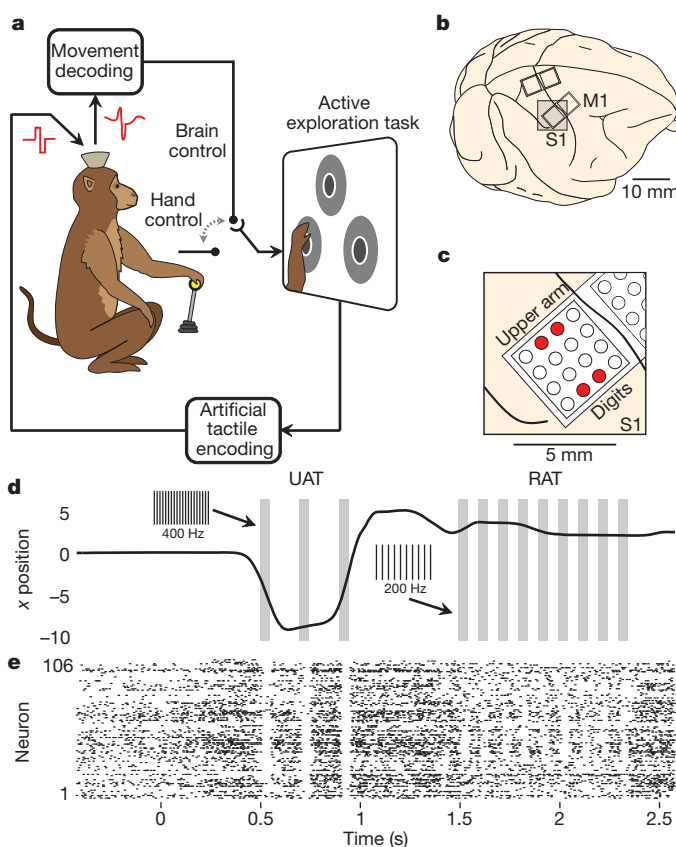


Figure 1 | The brain–machine–brain interface. **a**, Movement intentions are decoded from M1; artificial tactile feedback is delivered to S1. **b**, Microwires were implanted in M1 and S1. **c**, Microwires used for ICMS in monkey M are accentuated in red. **d**, Actuator movements for a trial in which monkey M explores UAT but ultimately selects RAT. Grey bars indicate stimulation patterns; insets indicate the ICMS frequency. **e**, Rastergram of M1 neurons recorded during the trial shown in **d**.

¹Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, USA. ²Center for Neuroengineering, Duke University, Durham, North Carolina 27710, USA. ³Department of Neurobiology, Duke University, Durham, North Carolina 27710, USA. ⁴STI IMT, École Polytechnique Fédérale de Lausanne, Lausanne CH1015, Switzerland. ⁵Department of Psychology and Neuroscience, Duke University, Durham, North Carolina 27708, USA. ⁶Edmond and Lily Safra International Institute of Neuroscience of Natal, Natal 59066-060, Brazil.

of microwires to the leg representation in monkey N. Each artificial texture consisted of a high-frequency pulse train presented in packets at a lower, secondary, frequency (Fig. 1d and Supplementary Fig. 2a). The rewarded artificial texture (RAT) consisted of 200-Hz pulse trains delivered in 10-Hz packets. The comparison artificial textures comprised 400-Hz pulse trains delivered in 5-Hz packets (unrewarded artificial texture (UAT)) or an absence of ICMS (null artificial texture (NAT)).

The main challenge solved here was the real-time coupling of ICMS feedback to the BMI decoder. Because ICMS artefacts masked neuronal activity for 5–10 ms after each pulse (Fig. 1d, e), we multiplexed neuronal recordings and ICMS with a 20-Hz clock rate (Supplementary Fig. 2a). The interleaved intervals proved adequate for online motor control and artificial sensation—a result that was not clear a priori because S1 stimulation could have affected M1 processing through the connections between these areas.

BMBI performance improved with training. In task I (Fig. 2a), monkey M surpassed chance performance after nine sessions and monkey N did so after four sessions ($P < 0.001$, one-sided binomial test). Improvement continued with more difficult tasks (tasks II–V) (Fig. 2a, b and Supplementary Fig. 3a). In particular, the time spent exploring unrewarded artificial textures decreased (Fig. 2c and Supplementary Fig. 3b). Additionally, performance improved over the course of daily experimental sessions (Fig. 2d). Psychometric analysis of RAT stimulation amplitudes (Supplementary Fig. 2b) indicated that at least 8 nC per ICMS waveform phase (100- μ s-wide current pulses of

80 μ A) was needed for the discrimination of artificial textures ($P < 0.001$, one-sided binomial test). Performance was at chance level for catch trials (task II), where ICMS was not delivered ($P = 0.90$, one-sided binomial test).

The statistics of object exploration intervals (total time spent over a particular object in a given trial) indicated that the monkeys uniquely discriminated each type of artificial texture (Figs 2c and 3a, c) and interpreted ICMS within hundreds of milliseconds—a timescale comparable to that for the discrimination of peripheral tactile stimuli^{24,25}. Early in task I, exploration intervals were equal for RAT and NAT ($P > 0.5$, Wilcoxon signed-rank test); with training, they became longer for RAT and shorter for NAT (tasks I and II) and UAT (tasks III–V). During hand control, the mean interval was longest for RAT (monkey M: $1,396 \pm 21$ ms; monkey N: $1,165 \pm 15$ ms; mean \pm s.e.m.), shortest for NAT (304 ± 8 ms; 300 ± 10 ms) and intermediate for UAT (452 ± 13 ms; 402 ± 14 ms) ($P < 0.01$, analysis of variance). During brain control, intervals spent exploring NAT (498 ± 15 ms; 587 ± 25 ms) and UAT (685 ± 20 ms; 764 ± 32 ms) were longer than they were during hand control, but were still shorter than those spent exploring RAT ($1,420 \pm 28$ ms; $1,398 \pm 55$ ms) ($P < 0.01$, analysis of variance).

Additional hallmarks of active exploration were seen in the conditional probabilities of selecting different artificial textures (Fig. 3b, d). During hand control trials, the monkeys stayed over the first-encountered artificial texture (arrows that loop back to the same

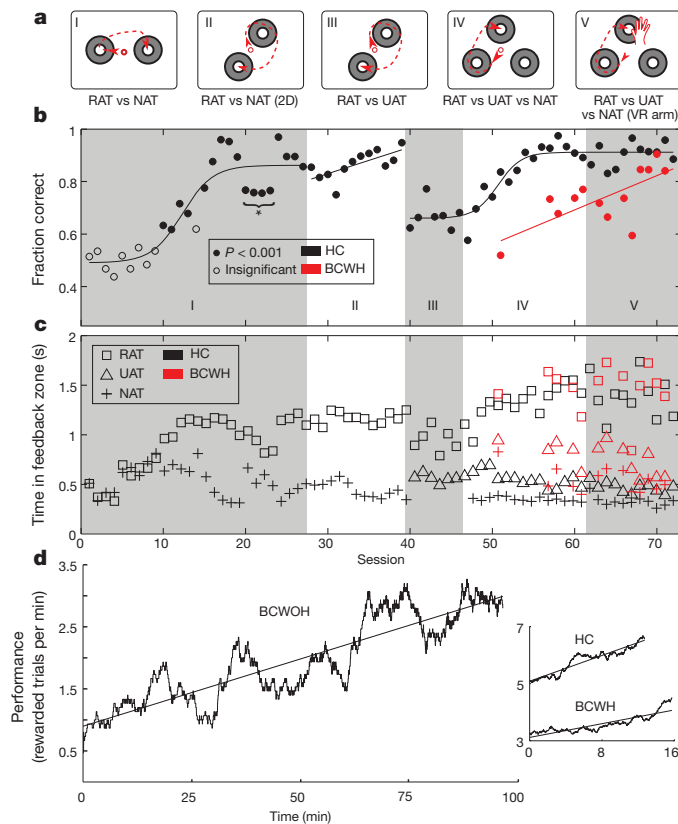


Figure 2 | Learning to use ICMS feedback. **a**, Behavioural tasks. 2D, two-dimensional; VR, virtual reality. **b**, **c**, Performance of monkey M (72 sessions). Circles (**b**) depict the fraction of correctly performed trials. Open circles indicate chance performance. Curves are lines of best fit. The asterisk indicates sessions used for psychometric measurements. Squares, triangles and crosses (**c**) represent mean times spent in RAT, UAT and NAT, respectively. Black, hand control; red, brain control. **d**, Intrasession performance for monkey M. Curves represent averages for brain control without hand movements (BCWOH; main panel, three sessions), for hand control (inset, 12 sessions) and for brain control with hand movements (BCWH; inset, 12 sessions). Lines are best linear fits.

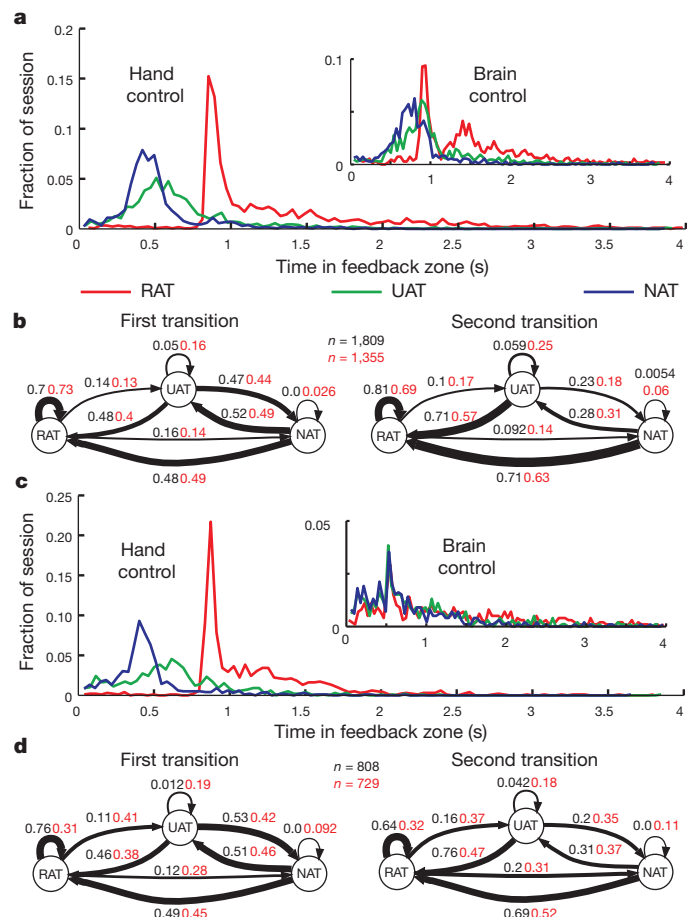


Figure 3 | Statistics of object exploration. **a**, Object exploration intervals during hand control and brain control (inset) for monkey M (hand control: $n = 1,809$ trials; brain control: $n = 1,355$ trials). **b**, State transition diagrams for monkey M, indicating the probabilities of reaching among RAT, UAT and NAT after the first (left subpanel) or second (right subpanel) reach. Black labels, hand control; red labels, brain control; line thickness is proportional to transition probability. **c**, Same as **a**, but for monkey N (hand control: $n = 808$ trials; brain control: $n = 729$ trials). **d**, Same as **b**, but for monkey N.

artificial texture in Fig. 3b, d) with high probability if it was RAT (monkey M: $P = 0.70$; monkey N: $P = 0.76$), but with low probability if it was UAT ($P = 0.05$; $P = 0.01$) or NAT ($P = 0.0$; $P = 0.0$) (Fig. 3b, d, left). After examining the second artificial texture, the monkeys could identify the correct artificial texture either by apprehending it directly or through a process of elimination. This follows from the increase from chance to approximately $P = 0.7$ in the probability of moving to RAT from NAT or UAT and the decrease to $P \approx 0.2$ in the probability of revisiting UAT or NAT (Fig. 3b, d, right). Similar effects were observed for brain control (Fig. 3b, d, red text).

Brain control started in task IV. During BCWH, the monkeys continued to hold the joystick although it was disconnected^{16,22}. During BCWOH^{9,22}, the joystick was removed. In monkey M, with more than 200 recorded neurons, performance was less accurate during BCWH ($73.75 \pm 3.00\%$; mean \pm s.e.m.) than during hand control ($91.48 \pm 1.20\%$). In monkey N, with 50 recorded neurons, performance dropped further ($50.37 \pm 3.74\%$ versus $91.45 \pm 1.91\%$), but still significantly exceeded the 33% chance level. M1 neurons showed directionally tuned modulations (Supplementary Figs 5 and 6) that were retained across different interfering ICMS patterns during both hand control (Supplementary Fig. 4a, b) and brain control (Fig. 4a, b).

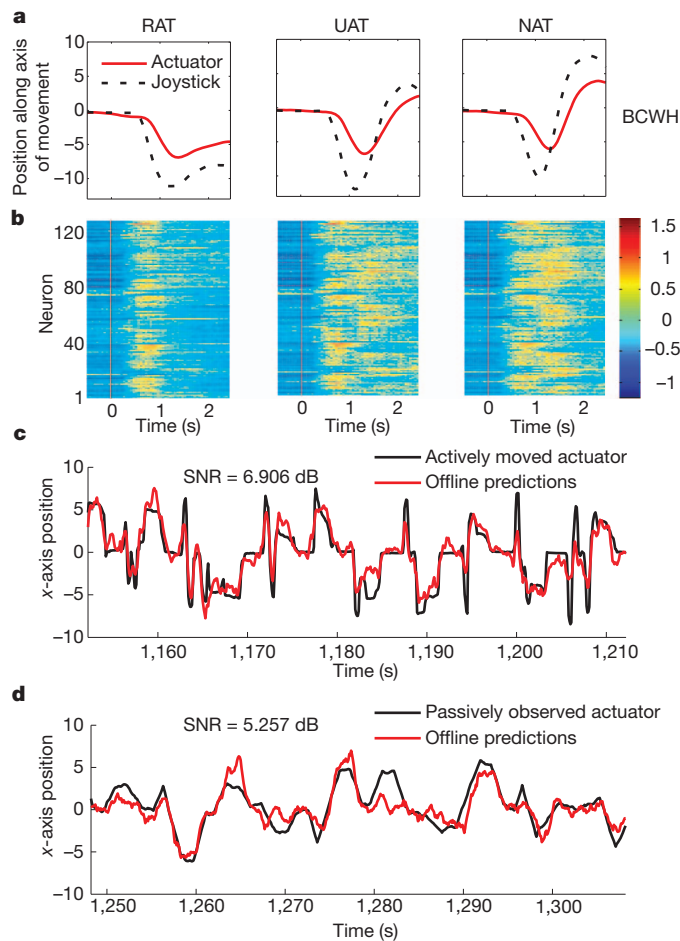


Figure 4 | M1 modulations during active control versus passive observation. **a, b,** Average brain control movements for monkey M ($n = 294$ trials) towards the RAT, UAT and NAT objects appearing on the left-hand side of the screen (**a**), and corresponding neuronal modulations (**b**). The colour scale shows normalized firing rate (Hz). Only trials with subsequent anticlockwise reaching movements are included in the middle and right subpanels. Red vertical lines indicate object onset. Firing rate normalized by s.d. **c, d,** Position of the actuator actively controlled (**c**) and passively observed (**d**) by monkey M. SNR, signal-to-noise ratio.

In BCWOH, task requirements were eased: the object selection period was reduced to 300–500 ms and monkeys were allowed to overstay at an incorrect object. The performance of monkey M, measured as the number of rewards per minute, steadily improved from 1.021 ± 0.007 to 2.962 ± 0.005 (mean \pm s.e.m.; Fig. 2d). Similar improvements were observed for hand control and BCWH (Fig. 2d, inset). The average frequency of actuator displacements, calculated from power spectra, was correlated with the improvement in performance during BCWOH ($R^2 = 0.16$ for the horizontal (x) coordinate and $R^2 = 0.26$ for the vertical (y) coordinate; $P < 0.001$, F -test), which indicated that the monkey modulated its brain activity to scan the targets faster. This behaviour was not random, as the exploration interval for NAT ($3,620 \pm 350$ ms; mean \pm s.e.m.) was significantly shorter ($P < 0.02$, Wilcoxon rank-sum test) than for UAT ($4,270 \pm 310$ ms). The exploration of RAT ($2,255 \pm 94$ ms) was the shortest owing to the reduced selection period. For monkey N, BCWOH performance (2.084 ± 0.085 rewards per minute) did not change within sessions, and the differences in exploration intervals were not significant.

In agreement with others^{26–30}, we observed that M1 neurons represented the movements of the actuator even when it was passively observed by the monkey (Supplementary Fig. 7). Actuator movements (task V) replayed for the monkeys could be reconstructed from M1 activity, using a separately trained decoder (Fig. 4d), with accuracy similar to that in reconstructions made for hand control (Fig. 4c). M1 representation of the passively viewed actuator is consistent with our suggestion that a neuroprosthetic limb might become incorporated in brain circuitry¹.

Our BMBI demonstrated direct bidirectional communication between a primate brain and an external actuator. Because both the afferent and efferent channels bypassed the subject's body, we propose that BMBIs can effectively liberate a brain from the physical constraints of the body. Accordingly, future BMBIs may not be limited to limb prostheses but may include devices designed for reciprocal communication among neural structures and with a variety of external actuators.

METHODS SUMMARY

All animal procedures were performed in accordance with the National Research Council's Guide for the Care and Use of Laboratory Animals and were approved by the Duke University Institutional Animal Care and Use Committee. Two rhesus monkeys were implanted with microwire arrays in both brain hemispheres. These implants were used for both recordings and ICMS (symmetric, biphasic, charge-balanced pulse trains; 100–200 μ s, 120–200 μ A). Monkeys manipulated a joystick to cause an actuator (computer cursor or a virtual-reality arm) to reach towards up to three objects displayed on a computer monitor. The task required searching for the single object with particular artificial tactile properties. Objects consisted of a central response zone and a peripheral feedback zone. Artificial tactile feedback was delivered when the actuator entered the feedback zone and continued in the response zone. Holding the actuator over the correct object for 0.8–1.3 s produced a reward (fruit juice). Holding the actuator over an incorrect object cancelled the trial. In brain control trials, the actuator was controlled by cortical ensemble activity decoded using an unscented Kalman filter²³. An interleaved scheme of alternating recording and stimulation subintervals (50 ms each, 50% duty cycle) was implemented to achieve concurrent afferent and efferent operations. In all offline analyses, ICMS periods were excluded from calculations of neuronal firing rates. The virtual-reality arm was animated using MOTIONBUILDER (Autodesk).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 June; accepted 17 August 2011.

Published online 5 October 2011.

1. Lebedev, M. A. & Nicolelis, M. A. Brain-machine interfaces: past, present and future. *Trends Neurosci.* **29**, 536–546 (2006).
2. Nicolelis, M. A. & Lebedev, M. A. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces. *Nature Rev. Neurosci.* **10**, 530–540 (2009).

3. Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. L. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci.* **2**, 664–670 (1999).
4. Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S. & Schwartz, A. B. Cortical control of a prosthetic arm for self-feeding. *Nature* **453**, 1098–1101 (2008).
5. Moritz, C. T., Perlmutter, S. I. & Fetz, E. E. Direct control of paralysed muscles by cortical neurons. *Nature* **456**, 639–642 (2008).
6. Wessberg, J. *et al.* Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408**, 361–365 (2000).
7. Taylor, D. M., Helms-Tillery, S. I. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**, 1829–1832 (2002).
8. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. Instant neural control of a movement signal. *Nature* **416**, 141–142 (2002).
9. Carmena, J. M. *et al.* Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1**, e42 (2003).
10. Johansson, R. S. & Westling, G. Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Exp. Brain Res.* **56**, 550–564 (1984).
11. Flanagan, J. R. & Wing, A. M. Modulation of grip force with load force during point-to-point arm movements. *Exp. Brain Res.* **95**, 131–143 (1993).
12. James, T. W., Kim, S. & Fisher, J. S. The neural basis of haptic object processing. *Can. J. Exp. Psychol.* **61**, 219–229 (2007).
13. Chatterjee, A., Aggarwal, V., Ramos, A., Acharya, S. & Thakor, N. V. A brain-computer interface with vibrotactile biofeedback for haptic information. *J. Neuroeng. Rehabil.* **4**, 40 (2007).
14. Kaczmarek, K., Webster, J., Bach-y-Rita, P. & Tompkins, W. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Trans. Biomed. Eng.* **38**, 1–16 (1991).
15. Marasco, P. D., Schultz, A. E. & Kuiken, T. A. Sensory capacity of reinnervated skin after redirection of amputated upper limb nerves to the chest. *Brain* **132**, 1441–1448 (2009).
16. O'Doherty, J. E., Lebedev, M. A., Hanson, T. L., Fitzsimmons, N. A. & Nicolelis, M. A. A brain-machine interface instructed by direct intracortical microstimulation. *Front. Integr. Neurosci.* **3**, 20 (2009).
17. Richer, F., Martinez, M., Robert, M., Bouvier, G. & Saint-Hilaire, J. M. Stimulation of human somatosensory cortex: tactile and body displacement perceptions in medial regions. *Exp. Brain Res.* **93**, 173–176 (1993).
18. London, B. M., Jordan, L. R., Jackson, C. R. & Miller, L. E. Electrical stimulation of the proprioceptive cortex (area 3a) used to instruct a behaving monkey. *IEEE Trans. Neural Syst. Rehabil. Eng.* **16**, 32–36 (2008).
19. Romo, R., Hernandez, A., Zainos, A. & Salinas, E. Somatosensory discrimination based on cortical microstimulation. *Nature* **392**, 387–390 (1998).
20. Fitzsimmons, N. A., Drake, W., Hanson, T. L., Lebedev, M. A. & Nicolelis, M. A. Primate reaching cued by multichannel spatiotemporal cortical microstimulation. *J. Neurosci.* **27**, 5593–5602 (2007).
21. Lederman, S. J. & Klatzky, R. L. Hand movements: a window into haptic object recognition. *Cognit. Psychol.* **19**, 342–368 (1987).
22. Lebedev, M. A. *et al.* Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. *J. Neurosci.* **25**, 4681–4693 (2005).
23. Li, Z. *et al.* Unscented Kalman filter for brain-machine interfaces. *PLoS ONE* **4**, e6243 (2009).
24. Lebedev, M. A., Denton, J. M. & Nelson, R. J. Vibration-entrained and premovement activity in monkey primary somatosensory cortex. *J. Neurophysiol.* **72**, 1654–1673 (1994).
25. Liu, Y., Denton, J. M. & Nelson, R. J. Neuronal activity in primary motor cortex differs when monkeys perform somatosensory and visually guided wrist movements. *Exp. Brain Res.* **167**, 571–586 (2005).
26. Cisek, P. & Kalaska, J. F. Neural correlates of mental rehearsal in dorsal premotor cortex. *Nature* **431**, 993–996 (2004).
27. Graziano, M. S., Cooke, D. F. & Taylor, C. S. Coding the location of the arm by sight. *Science* **290**, 1782–1786 (2000).
28. Maravita, A. & Iriki, A. Tools for the body (schema). *Trends Cogn. Sci.* **8**, 79–86 (2004).
29. Tkach, D., Reimer, J. & Hatsopoulos, N. G. Observation-based learning for brain-machine interfaces. *Curr. Opin. Neurobiol.* **18**, 589–594 (2008).
30. Dushanova, J. & Donoghue, J. Neurons in primary motor cortex engaged during action observation. *Eur. J. Neurosci.* **31**, 386–398 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Dimitrov for conducting the animal neurosurgeries; G. Lehw and J. Meloy for building brain implants; J. Fruh for rendering the virtual-reality monkey arm; T. Phillips, L. Oliveira and S. Halkiotis for technical support; and E. Thomson and Z. Li for comments. This research was supported by DARPA (award N66001-06-C-2019), TATRC (award W81XWH-08-2-0119), the NIH (award NS073125), NICHD/OD (award RC1HD063390) and NIH Director's Pioneer Award DP1OD006798, to M.A.L.N. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Office of the NIH Director or the NIH.

Author Contributions J.E.O'D., M.A.L. and M.A.L.N. designed experiments, analysed data and wrote the paper; J.E.O'D., M.A.L., P.J.I. and K.Z.Z. conducted experiments; and S.S. and H.B. developed the virtual-reality monkey arm.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A.L.N. (nicoleli@neuro.duke.edu).

METHODS

Subjects and implants. Two adult rhesus macaque monkeys (*Macaca mulatta*) participated in this study. Each monkey was implanted with four 96-microwire arrays constructed of stainless steel 304. Each hemisphere received two arrays: one in the upper-limb representation area and one in the lower-limb representation area. These arrays sampled neurons in both M1 and S1. We used recordings from the right-hemisphere arm arrays in each monkey, because each manipulated the joystick with its left hand. Within each array, microwires were grouped in two four-by-four, uniformly spaced grids each consisting of 16 electrode triplets. The separation between electrode triplets was 1 mm. The electrodes in each triplet had three different lengths, increasing in 300- μ m steps. The penetration depth of each triplet was adjusted with a miniature screw. After adjustments during the month following the implantation surgery, the depth of the triplets was fixed. The longest electrode in each triplet penetrated to a depth of 2 mm as measured from the cortical surface.

Tasks. The monkeys were trained to manipulate a computer cursor or a virtual-reality arm and to reach, using this actuator, towards objects displayed on a computer monitor. The objects were visually identical, but had different tactile properties as conveyed by ICMS of S1. In hand control, each trial commenced when the monkey held the joystick with their working hand. Then a target appeared in the centre of the screen. The monkey had to hold the actuator within that centre target for a random hold time uniformly drawn from the interval 0–2 s. After this, the central target disappeared and was replaced by a set of virtual objects radially arranged about the centre of the screen. Each of these consisted of a central response zone and a peripheral feedback zone, distinguished by their shading (Supplementary Fig. 1c). Tactile feedback was delivered in the feedback zone or the corresponding response zone. For monkey M, the radius of the response zone varied from 1.5 to 4.0 cm and the radius of the feedback zone varied from 4.5 to 7.25 cm, across all tasks and sessions. For monkey N, the radius of the response zone varied from 1.5 to 4.5 cm and the radius of the feedback zone varied from 4.75 to 9.5 cm, across all tasks and sessions. A trial was concluded when the monkey placed the actuator within the response zone for a hold interval (800–1,300 ms for hand control, depending on the session; 300–500 ms for brain control) or the monkey released the joystick handle (in hand control trials). The next trial commenced after an intertrial interval of 500 ms.

The sequence of events was the same during brain control trials. In some brain control sessions, the joystick was removed from the behavioural set-up. For these, each new trial commenced following the previous intertrial interval without the requirement for the monkey to hold the joystick. In tasks I–III, monkeys chose from a set of two objects. In task I, the monkeys had to choose between RAT and NAT for fixed object locations. In task II, RAT and NAT were presented on the screen at different angular locations in each trial. In task III, object number and spatial arrangement were the same as in task II, but RAT and UAT were used. In task IV, three objects were used (RAT, UAT and NAT) and their arrangement on the screen varied from trial to trial. Finally, in task V, the virtual-reality monkey arm replaced the computer cursor.

Psychometric measurements. Psychometric measurements determined the minimum ICMS amplitude that the monkeys could discriminate (Supplementary Fig. 2b). In these measurements, the ICMS amplitude was different in every trial. In each psychometric session, a range of amplitudes was selected such that about half were in a range clearly above the monkeys' threshold for discrimination and half were in a range of unknown discriminability.

Catch trials. In some sessions, a small percentage of trials (typically 1%) were designated as catch trials. In these trials, the microstimulator delivered pulse trains with zero amplitude, but all other aspects of the behavioural task remained the same. This allowed us to confirm that there were no unintentional sources of information that the monkeys could use to perform the tasks.

Algorithms. An Nth-order unscented Kalman filter²³ (UKF) was used for brain control predictions. Up to a tenth-order UKF was used in some sessions, but in most sessions we found that the third-order UKF was sufficient. The filter parameters were fitted on the basis of either the hand movements of the monkeys while they performed the task using a joystick or on passive observation of actuator movements while the monkeys' arms were restrained.

ICMS. Symmetric, biphasic, charge-balanced pulse trains were delivered in a bipolar fashion across pairs of microwires. The channels selected had clear sensory receptive fields in the upper limb (monkey M: two pairs of microwires with synchronous pulse trains) or lower limb (monkey N: one pair of microwires). For monkey M, the anodic and cathodic phases of stimulation had a pulse width of 105 μ s; for monkey N, the pulse width was 200 μ s. The anodic and cathodic phases of the stimulation waveforms were separated by 25 μ s.

Interleaved ICMS and recordings. We implemented an interleaved scheme of alternating recording and stimulation intervals (Supplementary Fig. 2a). Our BMI had a 10-Hz update rate. That is, 100 ms of past neural data were used to make predictions about the desired state of the actuator. We broke up each 100-ms interval into two 50-ms subintervals. In the first subinterval (Rec), neural activity was recorded as usual and the measured spike count was used to estimate the firing rate for the whole 100-ms interval. The second subinterval (Stim) was reserved exclusively for delivering ICMS; all spiking activity occurring in this subinterval was discarded. Whenever the actuator was in contact with a virtual object at the start of a Stim interval, an ICMS pulse train was delivered. For RAT, nine pulses of ICMS were delivered; for UAT, 18 pulses of ICMS were delivered; and for NAT, no pulses of ICMS were delivered. The neural activity in the Stim interval was discarded even for NAT, so that there would be no bias induced by ICMS-occluded neural data.

Virtual-reality monkey arm. In task V, we introduced a novel, brain-controlled virtual-reality arm with realistic kinematic movements and spatial interactions. The control loop rate was 50 Hz, with visual refreshing at 30 Hz. The arm model was designed to depict a rhesus macaque. We presented a first-person perspective of the virtual-reality arm to the monkey, who controlled the position of the hand. Arm posture was controlled using a mixture of direct control of end effectors and inverse kinematics, constrained by the physical interdependencies of the joints.

Clearance of p16^{Ink4a}-positive senescent cells delays ageing-associated disorders

Darren J. Baker^{1,2,3}, Tobias Wijshake^{1,4}, Tamar Tchkonja³, Nathan K. LeBrasseur^{3,5}, Bennett G. Childs¹, Bart van de Sluis⁴, James L. Kirkland³ & Jan M. van Deursen^{1,2,3}

Advanced age is the main risk factor for most chronic diseases and functional deficits in humans, but the fundamental mechanisms that drive ageing remain largely unknown, impeding the development of interventions that might delay or prevent age-related disorders and maximize healthy lifespan. Cellular senescence, which halts the proliferation of damaged or dysfunctional cells, is an important mechanism to constrain the malignant progression of tumour cells^{1,2}. Senescent cells accumulate in various tissues and organs with ageing³ and have been hypothesized to disrupt tissue structure and function because of the components they secrete^{4,5}. However, whether senescent cells are causally implicated in age-related dysfunction and whether their removal is beneficial has remained unknown. To address these fundamental questions, we made use of a biomarker for senescence, p16^{Ink4a}, to design a novel transgene, *INK-ATTAC*, for inducible elimination of p16^{Ink4a}-positive senescent cells upon administration of a drug. Here we show that in the *BubR1* progeroid mouse background, *INK-ATTAC* removes p16^{Ink4a}-positive senescent cells upon drug treatment. In tissues—such as adipose tissue, skeletal muscle and eye—in which p16^{Ink4a} contributes to the acquisition of age-related pathologies, life-long removal of p16^{Ink4a}-expressing cells delayed onset of these phenotypes. Furthermore, late-life clearance attenuated progression of already established age-related disorders. These data indicate that cellular senescence is causally implicated in generating age-related phenotypes and that removal of senescent cells can prevent or delay tissue dysfunction and extend healthspan.

To examine the role of cellular senescence in ageing and age-related pathologies, we designed a transgenic strategy for the clearance of senescent cells in mice. We based our approach on an earlier mouse model, termed FAT-ATTAC (fat apoptosis through targeted activation of caspase), in which adipocytes were selectively killed by apoptosis upon the administration of AP20187, a synthetic drug that induces dimerization of a membrane-bound myristoylated FK506-binding-protein–caspase 8 (FKBP–Casp8) fusion protein expressed specifically in adipocytes via the minimal *Fabp4* promoter⁶. Although a universal marker that is solely expressed in senescent cells has not been identified, most senescent cells seem to express p16^{Ink4a}, a cyclin-dependent kinase inhibitor and tumour suppressor that enforces growth arrest by activating Rb^{5,7}. Additionally, the expression of p16^{Ink4a} is known to increase with ageing in several rodent and human tissues⁸. We replaced the *Fabp4* promoter with a 2,617-bp fragment of the p16^{Ink4a} gene promoter that is transcriptionally active in senescent, but not non-senescent cells (Fig. 1a)⁹. We added an internal ribosome entry site (IRES) followed by an open reading frame (ORF) coding for enhanced green fluorescence protein (EGFP) to allow for detection and collection of p16^{Ink4a}-positive senescent cells. Injection of the resulting construct into fertilized eggs yielded nine transgenic *INK-ATTAC* founder lines.

To examine whether removal of p16^{Ink4a}-expressing cells is technically feasible and whether this affects age-associated deficits in mice,

we bred each of the founder lines onto a *BubR1* hypomorphic (*BubR1*^{H/H}) genetic background. *BubR1* encodes a key member of the mitotic checkpoint, a surveillance mechanism that ensures accurate chromosome segregation in mitosis by inhibiting the ubiquitin ligase activity of Cdc20-activated anaphase-promoting complex (APC^{Cdc20}) in the presence of unattached chromosomes^{10,11}. *BubR1*^{H/H} mice have a markedly shortened lifespan and exhibit a variety of age-related phenotypes, including infertility, lordokyphosis, sarcopenia, cataracts, fat loss, cardiac arrhythmias, arterial wall stiffening, impaired wound healing and dermal thinning^{12–14}. It has been proposed that *BubR1* is a determinant of natural ageing, because levels of *BubR1* decline markedly with age^{12–14}. *BubR1*^{H/H} mice selectively accumulate p16^{Ink4a}-positive cells in certain tissues in which age-associated pathologies develop, including adipose tissue, skeletal muscle and eye¹⁵. Inactivation of p16^{Ink4a} in these mice is known to delay the onset of age-related phenotypes selectively in these tissues¹⁵. To screen for *INK-ATTAC* transgene activity in p16^{Ink4a}-positive cells, we collected samples of inguinal adipose tissue (IAT) from each of the nine *BubR1*^{H/H}; *INK-ATTAC* strains at 5 months of age and analysed them for GFP expression by fluorescence microscopy. We observed GFP fluorescence in two of these strains, *BubR1*^{H/H}; *INK-ATTAC*-3 and -5 (Fig. 1b and Supplementary Fig. 1a). Quantitative reverse transcription–polymerase chain reaction (qRT–PCR) analysis of various tissues from *BubR1*^{H/H}; *INK-ATTAC*-3 and -5 mice demonstrated that *INK-ATTAC* and *GFP* transcript levels were significantly elevated in adipose tissue, skeletal muscle and eye, but not in tissues in which endogenous p16^{Ink4a} is not induced, including liver and heart (Fig. 1c and Supplementary Fig. 1b).

To confirm that transgenic *INK-ATTAC* and endogenous p16^{Ink4a} are under the same transcriptional control mechanism outside the context of *BubR1* hypomorphism, we harvested bone marrow cells from 2-month-old wild-type (WT); *INK-ATTAC*-3 and -5 mice and cultured them in the absence or presence of rosiglitazone, a drug that can induce cellular senescence and p16^{Ink4a} expression through activation of PPARγ¹⁶. Immunofluorescence microscopy revealed that a high proportion of cells expressed Flag-tagged FKBP–Casp8 in the presence of rosiglitazone, but not in its absence (Fig. 1d). Furthermore, we observed selective *INK-ATTAC* transgene induction in tissues of WT; *INK-ATTAC*-3 mice showing elevated expression of endogenous p16^{Ink4a} upon chronological ageing (Supplementary Fig. 2). Together, these data indicate that *INK-ATTAC* gene activity in founder lines 3 and 5 overlaps with endogenous p16^{Ink4a} expression.

Next, we tested whether *INK-ATTAC* is expressed in senescent cells in *BubR1* hypomorphic tissue. Fat tissue of aged *BubR1*^{H/H}; *INK-ATTAC* mice stained strongly for senescence-associated-β-galactosidase (SA-β-Gal; Fig. 1e). qRT–PCR analysis demonstrated that *INK-ATTAC* expression correlates with expression of senescence markers in IAT (Fig. 1f and Supplementary Fig. 3a). Skeletal muscle and lens tissue of aged *BubR1*^{H/H}; *INK-ATTAC* mice are SA-β-Gal negative (data not shown), but both these tissues expressed other markers of senescence

¹Department of Pediatric and Adolescent Medicine, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. ²Molecular Biology and Biochemistry, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. ³Robert and Arlene Kogod Center on Aging, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA. ⁴Department of Pathology and Medical Biology, University Medical Center Groningen, Groningen University, Groningen 9700 RB, The Netherlands. ⁵Physical Medicine and Rehabilitation, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA.

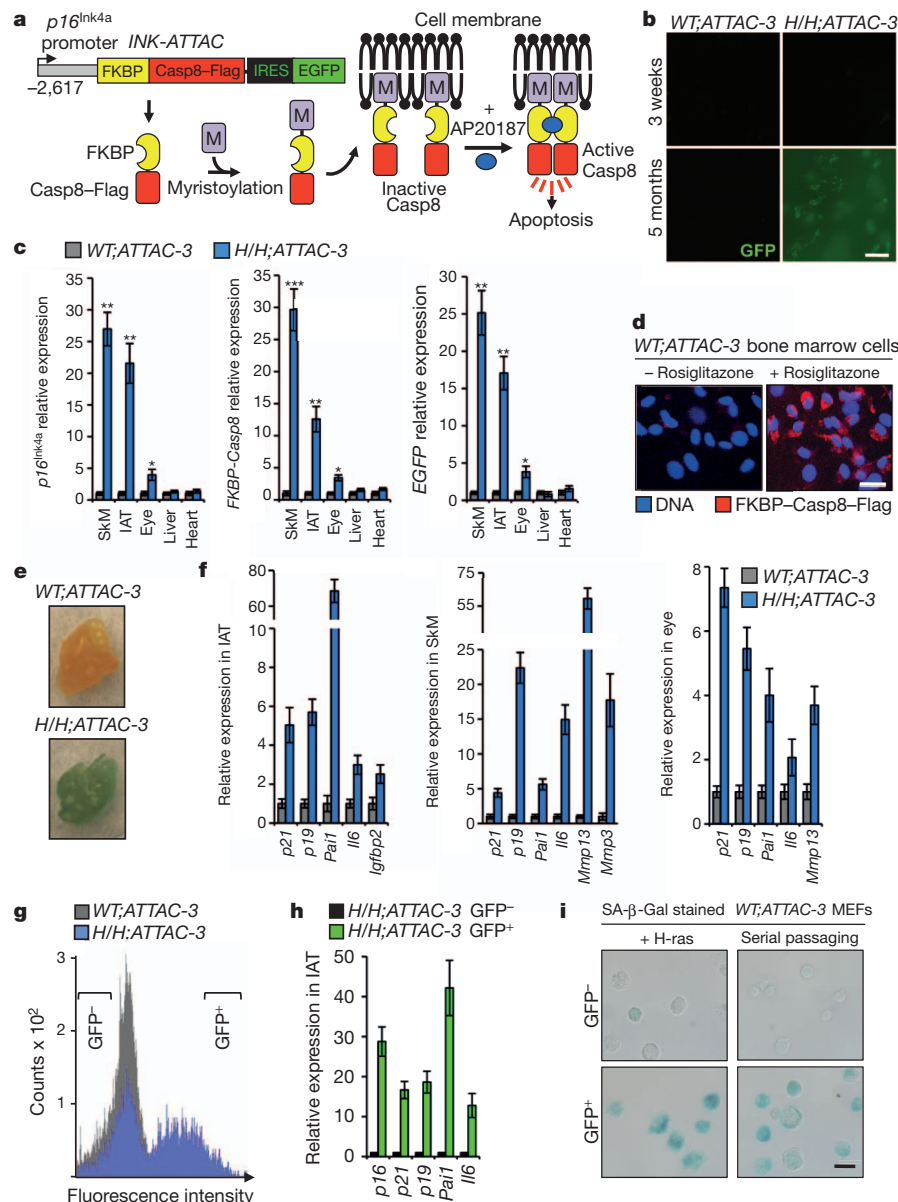


Figure 1 | Generation and characterization of *INK-ATTAC* transgenic mice. **a**, Schematic of the *INK-ATTAC* construct and the mechanism of apoptosis activation. **b**, GFP intensity of IAT. **c**, qRT-PCR analysis of the indicated tissues of 10-month-old mice. *ATTAC*, *INK-ATTAC*; *H/H*, *BubR1^{H/H}*; SkM, skeletal muscle (gastrocnemius). **d**, Bone marrow cells harvested from 2-month-old mice immunostained for Flag after culture in the absence or presence of rosiglitazone for 48 h. **e**, SA-β-Gal stained IAT collected from 9-month-old mice of the indicated genotypes. **f**, Expression of senescence markers in tissues of 10-month-old mice measured by qRT-PCR.

(Fig. 1f and Supplementary Fig. 3a). Senescence markers were not elevated in 3-week-old *BubR1^{H/H}*; *INK-ATTAC* mice (Supplementary Fig. 3b, c). To obtain additional evidence for selective expression of *INK-ATTAC* in senescent cells, we collected IAT from aged *BubR1^{H/H}*; *INK-ATTAC* animals, prepared single-cell suspensions by collagenase treatment, separated GFP⁺ and GFP⁻ cell populations by fluorescence activated cell sorting (FACS; Fig. 1g), and analysed each population for expression of *INK-ATTAC* and senescence markers by qRT-PCR. GFP⁺ cells not only expressed much higher levels of *p16^{Ink4a}* than GFP⁻ cells but also had elevated levels of other key senescence markers (Fig. 1h and Supplementary Fig. 3d). Furthermore, two conditions that induce *p16^{Ink4a}* expression and senescence in primary mouse embryonic fibroblasts (MEFs), ectopic expression of oncogenic Ras and serial passaging^{12,17,18}, produced a subpopulation of

All increases are statistically significant ($P < 0.05$). **g**, FACS profile of single-cell suspensions from IAT of 10-month-old mice. Brackets indicate sorting gates. **h**, GFP⁺ and GFP⁻ cell populations from IAT analysed for relative expression of senescence markers by qRT-PCR. All increases are statistically significant ($P < 0.01$). **i**, Bright field images of MEFs sorted into GFP⁺ and GFP⁻ populations after induction of senescence and then stained for SA-β-Gal. For all experiments, $n = 3$ untreated females per genotype. Error bars, s.d. Scale bars in **b**, **d** and **i**, 20 μ m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

GFP⁺ *WT;INK-ATTAC-3* MEFs that, in contrast to the remaining GFP⁻ cells, stained positively for SA-β-Gal (Fig. 1i). Taken together, these results indicate that *INK-ATTAC* is selectively expressed in *p16^{Ink4a}*-positive senescent cells.

To determine whether *INK-ATTAC* can eliminate senescent cells, we cultured bone marrow cells of *WT;INK-ATTAC* transgenic lines 3 and 5 in the presence of rosiglitazone to induce senescence and then monitored cell survival after activating the FKBP-Casp8 fusion protein by AP20187 treatment. We found that the vast majority of cells from both transgenic lines were either dead or in the process of dying 48 h after adding AP20187 (Fig. 2a). In contrast, parallel cultures that remained untreated consisted almost entirely of viable SA-β-Gal-positive cells. These data show that FKBP-Casp8 activation efficiently eliminates *p16^{Ink4a}*-positive senescent cells *in vitro*.

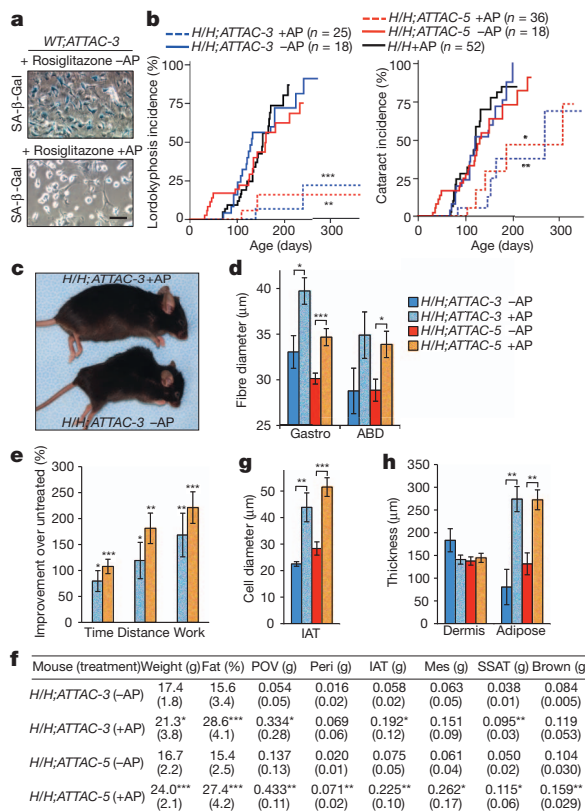
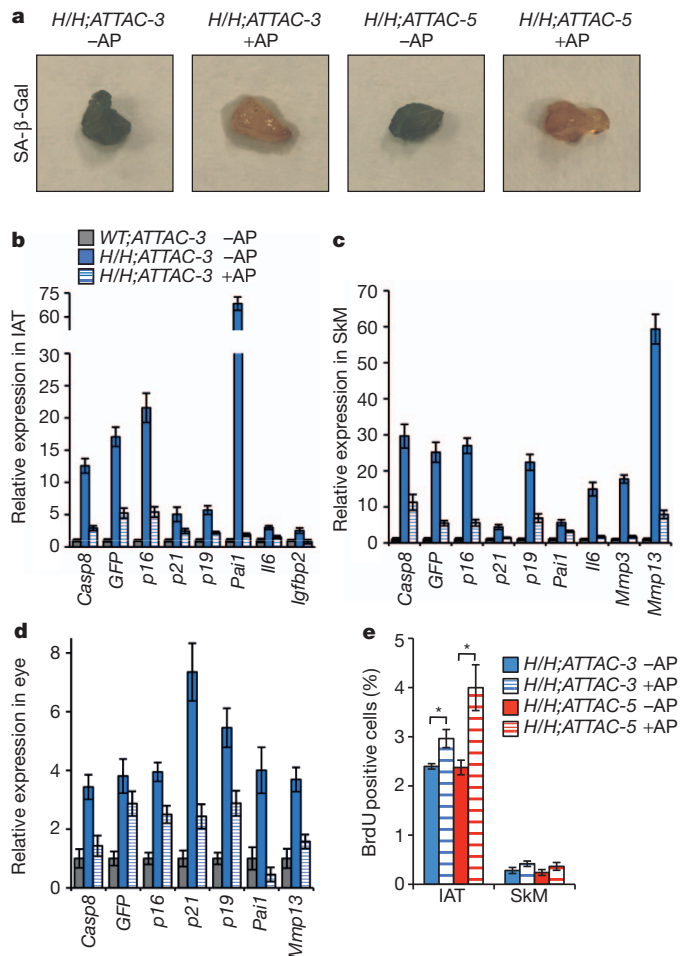


Figure 2 | *BubR1*^{H/H};INK-ATTAC mice treated with AP20187 from weaning age on show delayed onset of p16^{Ink4a}-mediated age-related phenotypes. **a**, Bone marrow cells cultured in rosiglitazone for 5 days and then treated or not treated with AP20187 (AP) for 2 days before SA-β-Gal staining. Scale bar, 50 μm. **b**, Incidence of lordokyphosis and cataracts. **c**, Representative images of 9-month-old mice. **d**, Mean skeletal muscle fibre diameters of 10-month-old mice. ABD, abdominal muscle; Gastro, gastrocnemius muscle. **e**, Exercise ability of 10-month-old AP20187-treated mice relative to age-matched untreated mice. Time is running time to exhaustion; distance is distance travelled at time of exhaustion; work is the energy expended to exhaustion. **f**, Body and fat depot weights of 10-month-old mice. Parentheses, s.d. Mes, mesenteric; Peri, perirenal; POV, paraovarian; SSAT, subscapular adipose tissue. **g**, Average fat cell diameters in IAT of 10-month-old mice. **h**, Dermis and subdermal adipose layer thickness of 10-month-old mice. Colour codes in **e**, **g** and **h** are as indicated in **d**. Error bars, s.e.m. For all analysis $n = 6$ female mice per genotype (per treatment). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Next, we examined whether clearance of p16^{Ink4a}-expressing cells from *BubR1*^{H/H} mice prevents or delays the onset of age-related phenotypes in this progeroid background. To this end, we established cohorts of *BubR1*^{H/H};INK-ATTAC-3 and -5 mice, which were either treated with AP20187 every third day beginning at 3 weeks of age or left untreated. Both treated and untreated mice were monitored for the development of age-associated deficits known to accompany p16^{Ink4a} induction, including sarcopenia, cataracts and loss of adipose tissue¹⁵. Remarkably, treated mice of both *BubR1*^{H/H};INK-ATTAC lines had substantially delayed onset of lordokyphosis (a measure of sarcopenia onset in this model¹⁵) and cataracts compared to untreated mice (Fig. 2b, c). Consistent with decreased lordokyphosis, muscle fibre diameters of AP20187-treated *BubR1*^{H/H};INK-ATTAC animals were larger than those of untreated counterparts (Fig. 2d). In addition to muscle retention, treadmill exercise tests revealed that duration of exercise, distance travelled and overall amount of work performed were all significantly increased in the animals treated with AP20187 (Fig. 2e), indicating preservation of muscle function. Dual-energy X-ray absorptiometry (DEXA) scans of *BubR1*^{H/H};INK-ATTAC mice confirmed that AP20187 treatment prevented loss of adipose tissue

(Fig. 2f). All major fat deposits were larger in AP20187-treated *BubR1*^{H/H};INK-ATTAC animals (Fig. 2f) and individual adipocytes were markedly increased in size (Fig. 2g). Consistent with this generally increased adiposity, lateral skin contained significantly more subdermal adipose tissue (Fig. 2h). The above age-related phenotypes were not delayed upon AP20187 treatment of *BubR1*^{H/H} mice lacking INK-ATTAC (Fig. 2b and Supplementary Fig. 4).

Age-related phenotypes of *BubR1*^{H/H} mice that arise in a p16^{Ink4a}-independent fashion, such as cardiac arrhythmias and arterial wall stiffening¹⁴, were not attenuated in AP20187-treated *BubR1*^{H/H};INK-ATTAC mice (Supplementary Fig. 5a, b). This correlated with lack of INK-ATTAC induction in heart and aorta (Fig. 1c and Supplementary Fig. 5c). Cardiac failure is presumably the main cause of death in *BubR1*^{H/H} mice (data not shown), which could explain why the overall survival of AP20187-treated *BubR1*^{H/H};INK-ATTAC mice was not substantially extended (Supplementary Fig. 5d). To examine whether clearance of p16^{Ink4a}-positive cells might have any overtly negative side effects, WT;INK-ATTAC mice were continuously treated with AP20187 until 8 months of age; however, no such effects were observed (data not shown). Taken together, these results indicate that continuous removal of p16^{Ink4a}-expressing cells from *BubR1*^{H/H};INK-ATTAC



mice selectively delays age-related phenotypes that depend on $p16^{\text{Ink4a}}$ induction.

Next, we determined whether the delayed onset of age-related pathologies coincided with a reduction in the number of senescent cells in these tissues. The IAT of AP20187-treated $BubR1^{\text{H/H}}; \text{INK-ATTAC}$ mice showed a marked decrease in SA- β -Gal staining compared with the IAT of untreated counterparts (Fig. 3a). Corresponding decreases in other senescence-associated markers were also observed, as well as expected reductions in INK-ATTAC and GFP (Fig. 3b and Supplementary Fig. 6a). Skeletal muscle and eye had a similar

reduction in senescence indicators (Fig. 3c, d and Supplementary Fig. 6b, c). BrdU incorporation was lower in IAT and muscle tissue of untreated than treated animals (Fig. 3e), supporting the contention that senescence-associated replicative arrest is decreased upon administration of AP20187 in $BubR1^{\text{H/H}}; \text{INK-ATTAC}$ transgenic animals. Together, these data indicate that senescent cells were cleared from tissues and that this delays acquisition of age-related dysfunction in $BubR1$ hypomorphic mice.

To investigate the effect of senescent cell clearance later in life when age-related phenotypes are apparent in $BubR1^{\text{H/H}}$ mice, we started AP20187 treatment of $BubR1^{\text{H/H}}; \text{INK-ATTAC}$ mice at 5 months instead of weaning age and measured $p16^{\text{Ink4a}}$ -dependent age-related phenotypes at 10 months. Cataracts had already fully matured by the onset of AP20187 treatment and remained unchanged (data not shown). Importantly, late-life treated animals had increased mean muscle fibre diameters and showed improved performance in treadmill exercise tests (Fig. 4a, b). Furthermore, most fat depots of these animals were enlarged and adipocyte cell size and subdermal adipose layer thickness were significantly increased (Fig. 4c–e). Senescence markers were substantially reduced in both fat and skeletal muscle of AP20187-treated animals (Fig. 4f, g and Supplementary Fig. 7). Analysis of 5-month-old untreated $BubR1^{\text{H/H}}; \text{INK-ATTAC-5}$ mice revealed that the observed improvements in skeletal muscle and fat of late-life treated 10-month-old $BubR1^{\text{H/H}}; \text{INK-ATTAC-5}$ mice reflect attenuated progression of age-related declines rather than a reversal of ageing (Fig. 4a, c–e). Thus, late-life clearance of $p16^{\text{Ink4a}}$ -positive senescent cells attenuates progression of age-related decline in $BubR1$ hypomorphic mice.

Whether and how cellular senescence is related to age-related diseases, frailty and dysfunction has been one of the major open questions in the biology of ageing and clinical geriatrics¹. Here we present a novel transgenic mouse model that allows for the inducible removal of $p16^{\text{Ink4a}}$ -positive senescent cells. Remarkably, even though transcriptional regulation of endogenous $p16^{\text{Ink4a}}$ expression is highly complex, involving various transcriptional activators/repressors, epigenetic mechanisms and antisense non-coding RNA^{19–22}, we find that expression of INK-ATTAC driven by a relatively small portion of the $p16^{\text{Ink4a}}$ promoter closely overlaps with that of endogenous $p16^{\text{Ink4a}}$. By breeding INK-ATTAC mice into a progeroid mouse genetic background, we show that both life-long and late-life clearance of the $p16^{\text{Ink4a}}$ -expressing senescent cells selectively delayed age-related pathologies in tissues that accumulate these cells. Furthermore, our data indicate that acquisition of the senescence-associated secretory phenotype (SASP), which enables cells to secrete a variety of growth factors, cytokines and proteases⁴, contributes to age-related tissue dysfunction. There were no overt side effects of senescent cell clearance in our model, even though it has been postulated that senescent cells enhance certain types of tissue repair^{23,24}. Our proof-of-principle experiments demonstrate that therapeutic interventions to clear senescent cells or block their effects may represent an avenue for treating or delaying age-related diseases and improving healthy human lifespan.

METHODS SUMMARY

Mouse strains and drug treatments. The INK-ATTAC transgenic construct was made as follows. The FKBP–Casp8 fragment was subcloned from the aP2–ATTAC transgenic construct⁶ (a gift from P. Scherer) and inserted into pBlueScriptII (Stratagene). A 2,617-bp segment of the murine $p16^{\text{Ink4a}}$ promoter was PCR amplified from BAC DNA to replace the aP2 promoter. An IRES–EGFP fragment was inserted 3' of the ATTAC. Transgenic founders were obtained by pronuclear injection of the INK-ATTAC construct into FVB oocytes. A PCR-based method was used for INK-ATTAC transgene identification (primer sequences are available upon request). $BubR1^{\text{H/H}}$ mice were generated as previously described¹². For AP20187 (ARIAD Pharmaceuticals) treatments, animals were injected intraperitoneally (i.p.) every 3 days with $0.2 \mu\text{g g}^{-1}$ body weight of the dimer-inducing drug⁶ from weaning ('life-long') or 5-months on ('late-life'). All mice were on a mixed 129 \times C57BL/6 \times FVB genetic background. Animals were housed in a pathogen-free barrier environment throughout the study. The Institutional Animal Care and Use Committee approved experimental procedures on mice.

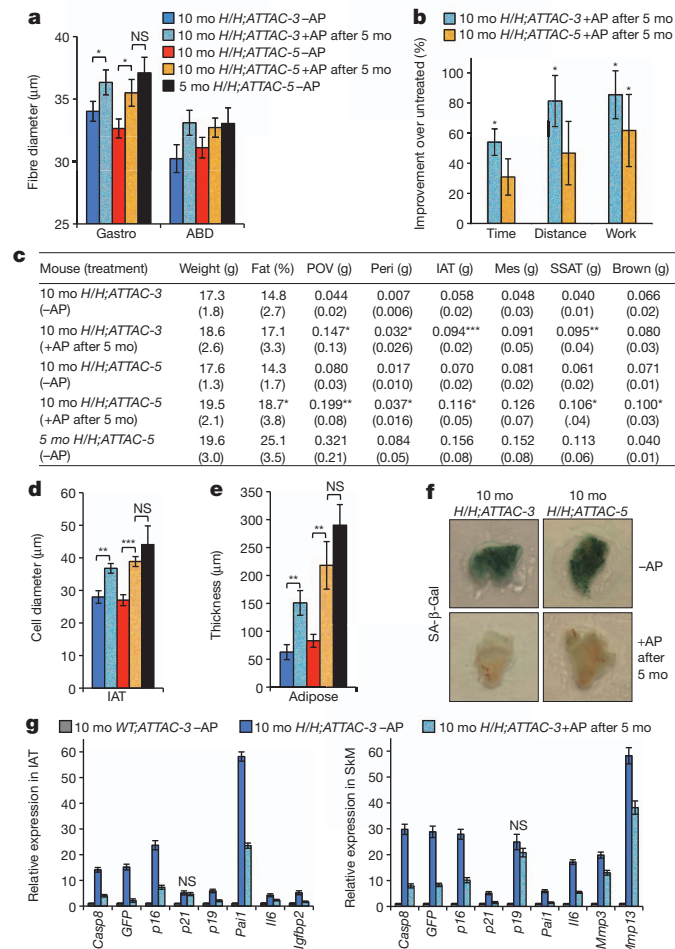


Figure 4 | Treatment of older $BubR1^{\text{H/H}}; \text{INK-ATTAC}$ mice with AP20187 delays progression of $p16^{\text{Ink4a}}$ -mediated age-related phenotypes. a, Mean skeletal muscle fibre diameters of the indicated mice. ABD, abdominal muscle; Gastro, gastrocnemius muscle. mo, months. b, Improvement of exercise ability of the indicated mice relative to age-matched untreated mice. c, Body and fat depot weights of the indicated mice. Parentheses, s.d. Mes, mesenteric; Peri, perirenal; POV, paraovarian; SSAT, subscapular adipose tissue. d, Average size of fat cells in IAT of the indicated mice. e, Subcutaneous adipose layer thickness of the indicated mice. f, SA- β -Gal-stained IAT. g, Expression of senescence markers in IAT and gastrocnemius of the indicated mice ($n = 3$ females per genotype per treatment). Expression of all genes, except those marked with NS, is significantly decreased ($P < 0.05$) upon late-life AP20187 treatment. Colour codes in d and e are as indicated in a. Error bars indicate s.e.m. except in g where they indicate s.d. For analyses in a–f $n = 5$ 5-month-old $BubR1^{\text{H/H}}; \text{INK-ATTAC-5}$ -AP females; $n = 9$ 10-month-old $BubR1^{\text{H/H}}; \text{INK-ATTAC-3}$ +AP and -AP females; $n = 7$ 10-month-old $BubR1^{\text{H/H}}; \text{INK-ATTAC-5}$ +AP females; and $n = 8$ 10-month-old $BubR1^{\text{H/H}}; \text{INK-ATTAC-5}$ -AP females.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. NS, not significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 May; accepted 30 September 2011.

Published online 2 November 2011.

- Campisi, J. Cellular senescence: putting the paradoxes in perspective. *Curr. Opin. Genet. Dev.* **21**, 107–112 (2011).
- Kuilman, T., Michaloglou, C., Mooi, W. J. & Peeper, D. S. The essence of senescence. *Genes Dev.* **24**, 2463–2479 (2010).
- Campisi, J. Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. *Cell* **120**, 513–522 (2005).
- Coppé, J. P. *et al.* Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. *PLoS Biol.* **6**, e301 (2008).
- Rodier, F. & Campisi, J. Four faces of cellular senescence. *J. Cell Biol.* **192**, 547–556 (2011).
- Pajvani, U. B. *et al.* Fat apoptosis through targeted activation of caspase 8: a new mouse model of inducible and reversible lipodystrophy. *Nature Med.* **11**, 797–803 (2005).
- Kim, W. Y. & Sharpless, N. E. The regulation of *INK4/ARF* in cancer and aging. *Cell* **127**, 265–275 (2006).
- Krishnamurthy, J. *et al.* *Ink4a/Arf* expression is a biomarker of aging. *J. Clin. Invest.* **114**, 1299–1307 (2004).
- Wang, W., Wu, J., Zhang, Z. & Tong, T. Characterization of regulatory elements on the promoter region of p16^{INK4a} that contribute to overexpression of p16 in senescent fibroblasts. *J. Biol. Chem.* **276**, 48655–48661 (2001).
- Malureanu, L. A. *et al.* BubR1 N terminus acts as a soluble inhibitor of cyclin B degradation by APC/C^{Cdc20} in interphase. *Dev. Cell* **16**, 118–131 (2009).
- Kulukian, A., Han, J. S. & Cleveland, D. W. Unattached kinetochores catalyze production of an anaphase inhibitor that requires a Mad2 template to prime Cdc20 for BubR1 binding. *Dev. Cell* **16**, 105–117 (2009).
- Baker, D. J. *et al.* BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nature Genet.* **36**, 744–749 (2004).
- Hartman, T. K., Wengenack, T. M., Poduslo, J. F. & van Deursen, J. M. Mutant mice with small amounts of BubR1 display accelerated age-related gliosis. *Neurobiol. Aging* **28**, 921–927 (2007).
- Matsumoto, T. *et al.* Aging-associated vascular phenotype in mutant mice with low levels of BubR1. *Stroke* **38**, 1050–1056 (2007).
- Baker, D. J. *et al.* Opposing roles for p16^{INK4a} and p19^{Arf} in senescence and ageing caused by BubR1 insufficiency. *Nature Cell Biol.* **10**, 825–836 (2008).
- Gan, Q. *et al.* PPAR γ accelerates cellular senescence by inducing p16^{INK4a} expression in human diploid fibroblasts. *J. Cell Sci.* **121**, 2235–2245 (2008).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).
- Kim, H. *et al.* Expression profiles of p53-, p16^{INK4a}-, and telomere-regulating genes in replicative senescent primary human, mouse, and chicken fibroblast cells. *Exp. Cell Res.* **272**, 199–208 (2002).
- Popov, N. & Gil, J. Epigenetic regulation of the *INK4b-ARF-INK4a* locus: in sickness and in health. *Epigenetics* **5**, 685–690 (2010).
- Gil, J. & Peters, G. Regulation of the *INK4b-ARF-INK4a* tumour suppressor locus: all for one or one for all. *Nature Rev. Mol. Cell Biol.* **7**, 667–677 (2006).
- Burd, C. E. *et al.* Expression of linear and novel circular forms of an *INK4/ARF*-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* **6**, e1001233 (2010).
- Li, J., Poi, M. J. & Tsai, M. D. Regulatory mechanisms of tumor suppressor P16^{INK4A} and their relevance to cancer. *Biochemistry* **50**, 5566–5582 (2011).
- Krizhanovsky, V. *et al.* Senescence of activated stellate cells limits liver fibrosis. *Cell* **134**, 657–667 (2008).
- Jun, J. I. & Lau, L. F. The matricellular protein CCN1 induces fibroblast senescence and restricts fibrosis in cutaneous wound healing. *Nature Cell Biol.* **12**, 676–685 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Zhou, D. Norris, T. Mann, U. Moedder, T. Pirtskhalava and S. Yamada for assistance; S. Khosla, T. von Zglinicki, L. Malureanu, R. Ricke and P. Galarzy, and members of the J.M.v.D. laboratory for helpful discussions; and P. Scherer for the gift of the aP2-ATTAC plasmid. This work was supported by the Ellison Medical Foundation (J.M.v.D.), the Noaber Foundation (J.M.v.D. and J.L.K.), the Robert and Arlene Kogod Center on Aging, and the National Institutes of Health (CA96985, J.M.v.D. and AG13925, J.L.K.).

Author Contributions D.J.B., T.T., J.L.K., and J.M.v.D. designed the *INK-ATTAC* strategy. D.J.B. and T.W. performed most of the experiments, T.T. did the rosiglitazone experiments, N.K.L. and B.G.C. assisted with the analysis of muscle functionality and *in vitro* senescence, respectively, and B.v.d.S. helped supervise T.W. The manuscript was written by D.J.B. and J.M.v.D. All authors discussed results, made figures and edited the manuscript. J.M.v.D. directed and supervised all aspects of the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.v.D. (vandeursen.jan@mayo.edu).

METHODS

Mouse strains and drug treatments. The *INK-ATTAC* transgenic construct was made as follows. The FKBP–Casp8 fragment was subcloned from the aP2–ATTAC transgenic construct⁶ (a gift from P. Scherer) and inserted into pBlueScriptII (Stratagene). A 2,617-bp segment of the murine *p16^{INK4a}* promoter was PCR amplified from BAC DNA to replace the aP2 promoter. An IRES–EGFP fragment was inserted 3' of the ATTAC. Transgenic founders were obtained by pronuclear injection of the *INK-ATTAC* construct into FVB oocytes. A PCR-based method was used for *INK-ATTAC* transgene identification (primer sequences are available upon request). *BubR1^{H/H}* mice were generated as previously described¹². For AP20187 (ARIAD Pharmaceuticals) treatments, animals were injected intraperitoneally (i.p.) every 3 days with 0.2 µg g⁻¹ body weight of the dimer-inducing drug⁸ from weaning ('life-long') or 5-months on ('late-life'). All mice were on a mixed 129 × C57BL/6 × FVB genetic background. Animals were housed in a pathogen-free barrier environment throughout the study. The Institutional Animal Care and Use Committee approved experimental procedures on mice.

Statistical analysis. Prism software was used for the generation of all survival curves and statistical analyses. Two-tailed unpaired *t* tests were used for pairwise significance analysis in the following figures: Fig. 1c, f and h; Fig. 2d–h; Fig. 3b–e; Fig. 4a–e, g; Supplementary Fig. 1b; Supplementary Fig. 2; Supplementary Fig. 3; Supplementary Fig. 4; Supplementary Fig. 5a–c; Supplementary Fig. 6; and Supplementary Fig. 7. Log-rank tests were used to determine overall and pairwise significance for incidence curves in Fig. 2b and survival curves in Supplementary Fig. 5d. For consistency in these comparisons, the following identifies the significance values: **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Cell culture. Culture of bone marrow cells was as previously described²⁵. Briefly, tibia and femur bones of 2-month-old *WT*;*INK-ATTAC* transgenic mouse lines were collected and flushed with DMEM containing 15% FBS. After centrifugation at 400g for 10 min and counting of viable cells with trypan blue, cells were resuspended in DMEM containing 15% FBS to a final concentration of 5 × 10⁶ viable cells per ml. Initially, cells were plated in 6-well tissue culture dishes at 3.5 ml well⁻¹ (1.9 × 10⁶ cells cm⁻²). Cultures were kept in a humidified 5% CO₂ incubator at 37 °C for 72 h, when non-adherent cells were removed by changing the medium. Assays were performed on cells that had been trypsinized and seeded to confluency in 24-well plates. To induce senescence and evaluate expression of the *INK-ATTAC* transgene, cells were treated with 1 µM rosiglitazone (Cayman Chemical Company) or with vehicle. The accumulation of GFP⁺ cells was observed by fluorescence microscopy and transgene expression was verified by immunofluorescence staining for Flag (Origene) as described²⁶. After 5 days of rosiglitazone treatment, cells were washed with PBS and treated with vehicle, 1 µM rosiglitazone, 10 nM AP20187, or both. After 48 h, cultures were fixed and stained for SA-β-Gal activity as described²⁷. *WT*;*INK-ATTAC* MEFs were generated and cultured as previously described¹². For induction of replicative senescence, *WT*;*INK-ATTAC* MEF cultures were maintained in 20% O₂ for 12–15 passages. For oncogene-induced senescence, early passage MEFs were infected with concentrated pBABE puro H-ras^{G12V} retrovirus (Addgene plasmid 9051) for 48 h. MEFs were then cultured in DMEM containing puromycin (2 µg ml⁻¹) for 5 days. Cells from serial passage and H-ras induced senescence were sorted into GFP⁺ and GFP⁻ populations using a FACS Aria Cell Sorter (BD Biosciences) running FACSDiva software (serial passaging and H-ras expression yielded cultures with approximately 90% and 50% GFP⁺ cells, respectively). Sorted cells were transferred to polyethylenimine-coated chambered slides and stained for SA-β-Gal according to manufacturer's instructions (Cell Signaling).

qRT–PCR and flow cytometry. RNA extraction, cDNA synthesis and qRT–PCR from whole-mouse tissue were performed as previously described¹⁵. To perform qRT–PCR on GFP⁺ and GFP⁻ cell populations of IAT, single-cell suspensions of stromal vascular fraction were prepared from ~50 mg IAT as described²⁸. Cell sorting was performed as described above. RNA was extracted from the collected cells using an RNeasy Micro Kit (Qiagen) and cDNA synthesized using a WT–Ovation RNA Amplification kit (NuGEN Technologies), according to the manufacturers' protocols. qRT–PCR primers were as follows: FKBP–Casp8 forward, GAATCACAGACTTTGGACAAAGTT; FKBP–Casp8 reverse, GGTCAAAGC CCCTGCATCCAAG; EGFP forward, CAAACTACAACAGCCACAA CG; EGFP reverse, GGTCACGAACCTCCAGCAG. Sequences of other primers used were as previously described¹⁵.

Analysis of progeroid phenotypes. Bi-weekly checks for lordokyphosis and cataracts were performed as described¹⁵. Skeletal muscle fibre diameter measurements were performed on cross-sections of gastrocnemius and abdominal muscles of female mice as described¹⁵. Fifty total fibres per sample were measured using a calibrated computer program (Olympus MicroSuite Five). Fat cell diameter measurements were performed on IAT according to the same method. Dissection, histology and measurements of dermal and adipose layers of skin were performed as described previously¹², although the lateral skin between the front and hind limb was used because this adipose layer is nearly three times thicker than dorsal skin. Measurements of body weight, length, gastrocnemius muscle and assorted adipose depots were performed on 10-month-old females. Bone mineral content, bone mineral density and total body adipose tissue were analysed by DEXA scanning as previously described⁶. Exercise measurements were performed on 10-month-old mice as previously described²⁹. Animals were acclimated for 3 days for 5 min at a speed of 5 m min⁻¹ before experimentation. For the experiment, the speed of the treadmill began at 5 m min⁻¹ and was increased to 8 m min⁻¹ after 2 min. Thereafter, the speed was increased at a rate of 2 m min⁻¹ every 2 min and the time (in seconds) and distance (in metres) to exhaustion, as defined by an inability to move along the treadmill with stimulation, were determined. The formula to determine the amount of work (*J*) performed was: mass (kg) × *g* (9.8 m s⁻²) × distance (m) × sin(*θ*) (with an incline of *θ* = 5°). Cardiac arrhythmia measurements were performed using a Vevo2100 ultrasound system (Visualsonics) as previously described³⁰.

In vivo BrdU incorporation and SA-β-Gal staining. Analyses for *in vivo* BrdU incorporation were performed in 10-month-old female mice as described¹⁵. Adipose tissue depots were stained for SA-β-Gal activity as previously described¹².

25. Soleimani, M. & Nadri, S. A protocol for isolation and culture of mesenchymal stem cells from mouse bone marrow. *Nature Protocols* **4**, 102–106 (2009).
26. Malureanu, L. et al. Cdc20 hypomorphic mice fail to counteract de novo synthesis of cyclin B1 in mitosis. *J. Cell Biol.* **191**, 313–329 (2010).
27. Dimri, G. P. et al. A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
28. Kirkland, J. L., Hollenberg, C. H. & Gillon, W. S. Effects of fat depot site on differentiation-dependent gene expression in rat preadipocytes. *Int. J. Obes. Relat. Metab. Disord.* **20** (Suppl 3), S102–S107 (1996).
29. LeBrasseur, N. K. et al. Myostatin inhibition enhances the effects of exercise on performance and metabolic outcomes in aged mice. *J. Gerontol. A Biol. Sci. Med. Sci.* **64A**, 940–948 (2009).
30. Martinez-Fernandez, A. et al. iPS programmed without c-MYC yield proficient cardiogenesis for functional heart chimerism. *Circ. Res.* **105**, 648–656 (2009).

ATM controls meiotic double-strand-break formation

Julian Lange¹, Jing Pan^{1†}, Francesca Cole², Michael P. Thelen³, Maria Jasin² & Scott Keeney^{1,4}

In many organisms, developmentally programmed double-strand breaks (DSBs) formed by the SPO11 transesterase initiate meiotic recombination, which promotes pairing and segregation of homologous chromosomes¹. Because every chromosome must receive a minimum number of DSBs, attention has focused on factors that support DSB formation². However, improperly repaired DSBs can cause meiotic arrest or mutation^{3,4}; thus, having too many DSBs is probably as deleterious as having too few. Only a small fraction of SPO11 protein ever makes a DSB in yeast or mouse⁵ and SPO11 and its accessory factors remain abundant long after most DSB formation ceases¹, implying the existence of mechanisms that restrain SPO11 activity to limit DSB numbers. Here we report that the number of meiotic DSBs in mouse is controlled by ATM, a kinase activated by DNA damage to trigger checkpoint signalling and promote DSB repair. Levels of SPO11–oligonucleotide complexes, by-products of meiotic DSB formation, are elevated at least tenfold in spermatocytes lacking ATM. Moreover, *Atm* mutation renders SPO11–oligonucleotide levels sensitive to genetic manipulations that modulate SPO11 protein levels. We propose that ATM restrains SPO11 via a negative feedback loop in which kinase activation by DSBs suppresses further DSB formation. Our findings explain previously puzzling phenotypes of *Atm*-null mice and provide a molecular basis for the gonadal dysgenesis observed in ataxia telangiectasia, the human syndrome caused by ATM deficiency.

SPO11 creates DSBs via a covalent protein–DNA intermediate that is endonucleolytically cleaved to release SPO11 attached to a short oligonucleotide, freeing DSB ends for further processing and recombination⁵ (Fig. 1a). SPO11–oligonucleotide complexes are a quantitative by-product of DSB formation that can be exploited to study DSB number and distribution^{5–7} (Supplementary Fig. 1). We examined SPO11–oligonucleotide complexes by SPO11 immunoprecipitation and 3'-end labelling of whole-testis extracts from *Atm*^{−/−} mutant mice, which have multiple catastrophic meiotic defects, including chromosome synapsis failure and apoptosis^{8–12}. The *Atm*^{−/−} phenotype resembles that of mutants lacking DSB repair factors such as DMCL1, indicating that absence of ATM causes meiotic recombination defects. Although *Spo11*^{−/−} mutation is epistatic to *Atm*^{−/−} (refs 11, 12), the functional relationship between ATM and SPO11 is complex, as meiotic defects of *Atm*^{−/−} mice are substantially rescued by reducing *Spo11* gene dosage^{13,14} (discussed later).

Unexpectedly, we found that adult *Atm*^{−/−} testes exhibited an approximately tenfold elevation in steady-state levels of SPO11–oligonucleotide complexes relative to wild-type littermates (Fig. 1b) (11.3 ± 4.5-fold, mean and standard deviation, *n* = 7 littermate pairs). This finding contrasts with *Dmcl1*^{−/−} testes, which showed a ~50% reduction in SPO11–oligonucleotide complexes (0.51 ± 0.06-fold relative to wild type, *n* = 5) (Fig. 1c), as previously shown^{5,7}. The mutants share similar arrest points in prophase I, as determined by molecular and histological data¹²; thus, increased SPO11–oligonucleotide complexes in *Atm*^{−/−} spermatocytes are not an indirect consequence

of arrest or of an increased fraction of meiocytes harbouring such complexes.

In *Atm*^{−/−} testes, levels of free SPO11 (that is, not bound to an oligonucleotide) were much lower than in wild type (Fig. 1b). This is not because a large fraction of SPO11 has been consumed in covalent complexes with DNA—which alters its electrophoretic mobility—as free SPO11 was not restored to wild-type levels by nuclease treatment (Fig. 1d). Instead, because *Spo11* transcript levels in wild type are highest in later stages of meiotic prophase^{15–18}, after the arrest point of *Atm*^{−/−} cells, reduced free SPO11 is attributable to the lack of later meiotic cell types, consistent with the reduced free SPO11 also found in

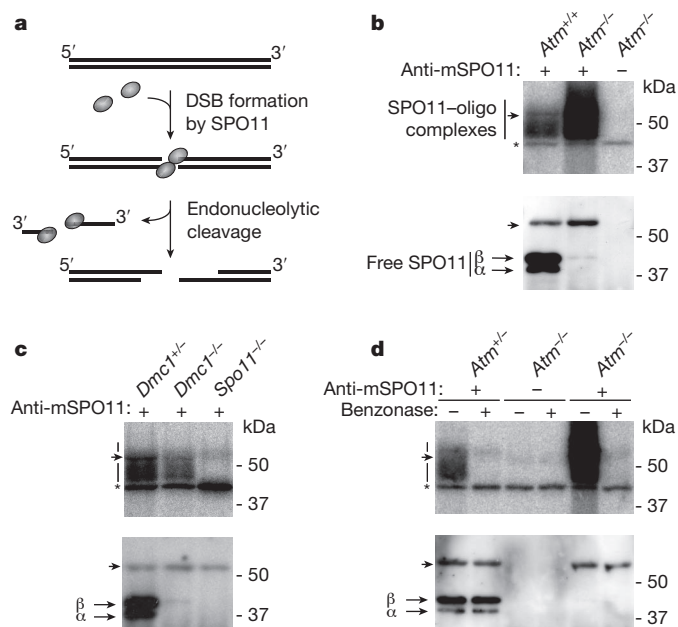


Figure 1 | SPO11 activity and expression in the absence of ATM. **a**, SPO11 attacks the DNA phosphodiester backbone, forming a covalent intermediate with the 5' strand termini of the DSB. Endonucleolytic cleavage removes SPO11 covalently attached to an oligonucleotide. **b**, **c**, Steady-state levels of SPO11–oligonucleotide (SPO11–oligo) complexes are elevated in *Atm*^{−/−} testes (**b**), but are decreased in *Dmcl1*^{−/−} testes (**c**). Anti-mSPO11, anti-mouse SPO11 antibody. SPO11 immunoprecipitates from extracts of whole adult testes were treated with terminal transferase and [α -³²P] dCTP, resolved by SDS–PAGE, and transferred to a membrane. Representative experiments using littermates of the indicated genotypes are shown. Top, autoradiograph. Bottom, anti-SPO11 western blot detection. Vertical lines, extent of SPO11-specific signals; α and β , major SPO11 isoforms; asterisk, non-specific terminal transferase labelling; arrowheads, migration position of immunoglobulin heavy chain. **d**, Treatment of labelled SPO11 immunoprecipitates with benzonase does not detectably alter levels of free SPO11, but this sequence non-specific nuclease efficiently removes the 3'-end label (compare lanes \pm benzonase), and was previously shown to completely remove DNA covalently bound to yeast Spo11 (ref. 1).

¹Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ²Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. ³Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, USA. ⁴Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [†]Present address: Cell Biology Department, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

Dmcl^{-/-} cells (Fig. 1c). As expected, the residual SPO11 protein in *Atm*^{-/-}, like *Dmcl*^{-/-}, testes was mostly SPO11 β (Fig. 1b, c). SPO11 α and SPO11 β are major protein isoforms encoded by developmentally regulated splice variants; SPO11 β is expressed earlier and is sufficient for nearly normal DSB levels^{5,15,17–20}.

Elevated SPO11–oligonucleotide complexes can be explained by an increased number of meiotic DSBs and/or a longer lifespan of complexes. To distinguish between these possibilities, we examined the initial appearance and persistence of SPO11–oligonucleotide complexes in juvenile mice, in which the first suite of spermatogenic cells proceeds through meiosis in a semi-synchronous fashion²¹. First, we assayed SPO11–oligonucleotide complexes in whole-testis extracts from wild-type C57BL/6J mice at postnatal days (d)7 to 24 (Fig. 2a). SPO11–oligonucleotide complexes first appeared between d9 and d10, when most cells of the initial cohort had entered leptotema. SPO11–oligonucleotide complexes persisted or increased slightly until d15, when the first cohort had progressed into pachynema. Levels rose still further from d16 to d18, coincident with the second cohort of spermatogenic cells reaching leptotema²¹. Thus, SPO11–oligonucleotide complexes appear at the same time as cell types that experience the majority of meiotic DSBs. Consistent with findings in mutants (see earlier), only trace amounts of free SPO11 protein were seen when SPO11–oligonucleotide complexes first appeared, with SPO11 β the predominant isoform at these times (Fig. 2a). Importantly, SPO11–oligonucleotide complex levels did not decline between the first and second spermatogenic cohorts. We infer that the lifespan of the complexes is long relative to the duration of prophase, and that an increased lifespan is not a likely explanation for the large increase in steady-state SPO11–oligonucleotides in adult *Atm*^{-/-} testes.

In support of this interpretation, we found that SPO11–oligonucleotide complexes were undetectable in *Atm*^{-/-} testes at d7 (data not shown) but were already elevated 3.3-fold compared with a wild-type littermate

when they first appeared, increasing to 8.4-fold over wild type by d12 (Fig. 2b). Because *Atm*^{-/-} juveniles showed higher SPO11–oligonucleotide levels as soon as the first leptotene cells appeared, we conclude that most, if not all, of the increase reflects a greater number of meiotic DSBs occurring during prophase I.

Meiotic defects of mice lacking ATM are substantially suppressed by reducing *Spo11* gene dosage: *Spo11*^{+/-} *Atm*^{-/-} spermatocytes pair and recombine their autosomes and progress through meiotic prophase to metaphase I, where they arrest due to a failure in sex chromosome pairing and recombination^{13,14}. The reason for this puzzling rescue was unknown, but our current findings suggest an explanation: the majority of meiotic defects in *Atm*-null spermatocytes are caused by grossly elevated DSB levels, which are lowered by *Spo11* heterozygosity (which reduces SPO11 protein levels by half in adult and juvenile testes (ref. 17 and our unpublished data)). Indeed, we found SPO11–oligonucleotide complexes in *Spo11*^{+/-} *Atm*^{-/-} mice to be substantially reduced compared with *Atm*^{-/-} littermates (Fig. 3a). The remaining increase in SPO11–oligonucleotide complexes in *Spo11*^{+/-} *Atm*^{-/-} mutants compared with wild type (range of 4.5- to 7.8-fold, *n* = 2) is not simply a consequence of metaphase arrest, because SPO11–oligonucleotide complexes were not elevated in mice that exhibit a similar arrest point due to absence of MLH1, a protein involved late in recombination²² (Fig. 3a). The fact that DSBs are still elevated in *Spo11*^{+/-} *Atm*^{-/-} spermatocytes relative to wild type may account for some or all of the remaining defects in this mutant, including axis interruptions at sites of ongoing recombination and persistent unrepaired DSBs late in prophase I (ref. 14).

Our findings indicate that the absence of ATM renders the extent of DSB formation sensitive to SPO11 expression levels. Therefore, we reasoned that increasing SPO11 expression should further elevate DSB formation in ATM-deficient cells. To test this prediction, we used a previously described transgene (*Xmr-Spo11 β*) that expresses the SPO11 β isoform¹⁸. Indeed, there was substantial further elevation of SPO11–oligonucleotide complex levels (20.9 ± 1.5 -fold over wild-type littermates, *n* = 3) upon introduction of this transgene in an *Atm*-null background with intact endogenous *Spo11* (Fig. 3b). By contrast, the transgene resulted in only a modest increase in SPO11–oligonucleotide complexes in an ATM-proficient background (1.1 ± 0.05 -fold, *n* = 3) (Fig. 3b).

SPO11–oligonucleotide complexes from *Atm*-null testes were consistently shifted to a higher electrophoretic mobility compared to wild type or other mutants (Figs 1, 2b and 3). To examine the distribution of

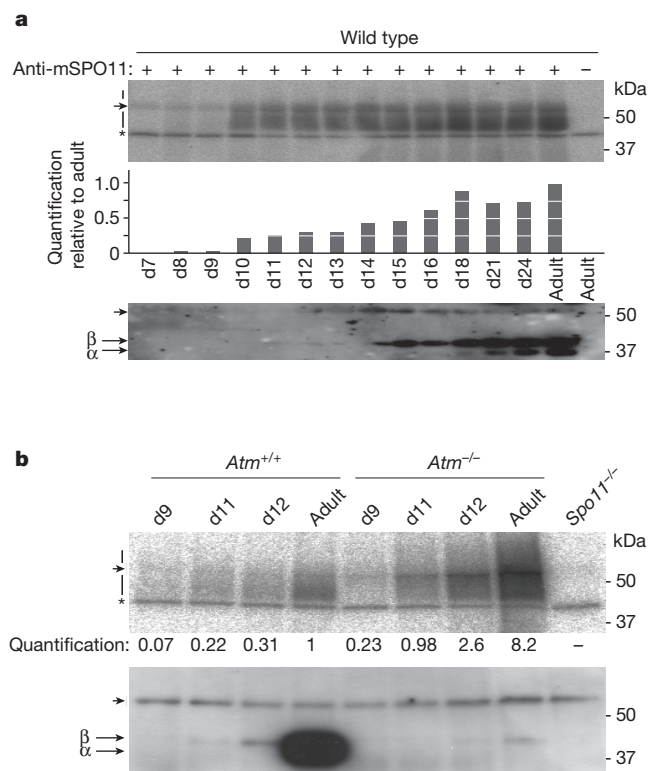


Figure 2 | SPO11–oligonucleotide complexes from juvenile mice.

a, SPO11–oligonucleotide complexes from testes of wild-type mice from d7–d24. Top, autoradiograph. Middle, quantification. Bottom, anti-SPO11 western detection. **b**, SPO11–oligonucleotides are elevated in testes from juvenile *Atm*^{-/-} mice. Top, autoradiograph. Bottom, anti-SPO11 western detection.

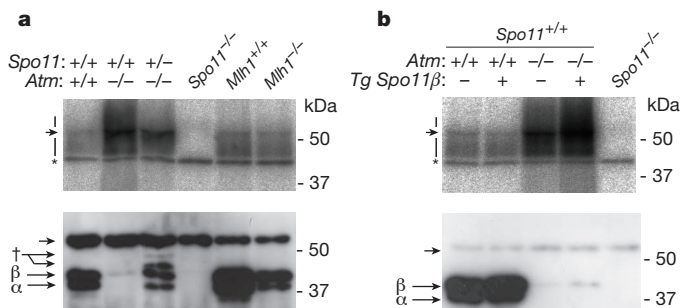


Figure 3 | *Spo11* gene dosage modulates SPO11–oligonucleotide complex levels in *Atm*-deficient spermatocytes. **a**, SPO11–oligonucleotide complexes are reduced in *Spo11*^{+/-} *Atm*^{-/-} testes relative to *Atm*^{-/-}, but are more abundant than in wild type or in an *Mlh1*^{-/-} mutant, which similarly arrests at metaphase. Consistent with further meiotic progression than *Atm*^{-/-}, both SPO11 isoforms (α and β) are expressed in *Spo11*^{+/-} *Atm*^{-/-} testes, although at reduced levels due to *Spo11* heterozygosity. Dagger, lower-mobility polypeptides probably originating from the *Spo11* knockout allele¹⁸. **b**, SPO11–oligonucleotide complexes are further elevated by SPO11 β expression from the *Xmr-Spo11 β* transgene (*Tg Spo11 β*)¹⁸ in *Atm*^{-/-} spermatocytes. Introducing this transgene into an otherwise wild-type background only modestly increased SPO11–oligonucleotide complex levels.

oligonucleotide lengths, labelled complexes were protease-digested and the resulting oligonucleotides were electrophoresed on a high-resolution gel (Fig. 4a). As previously shown⁵, SPO11-oligonucleotides from wild type have a bimodal length distribution with prominent subpopulations at apparent sizes of ~15–27 and ~31–35 nucleotides. *Atm*^{-/-} mice showed a different pattern with or without the *Spo11* transgene: oligonucleotides in the shorter size range were less abundant relative to the ~31–35 nucleotide class and longer oligonucleotides appeared, including an abundant class of ~40–70 nucleotides and a subpopulation that ranged to >300 nucleotides. *Spo11*^{+/-} *Atm*^{-/-} mice showed an intermediate pattern, with more pronounced enrichment of the ~31–35 nucleotide class relative to both smaller and longer oligonucleotides. These results indicate that ATM influences an early step in nucleolytic processing of meiotic DSBs, as has been proposed in yeast²³. In principle, altered oligonucleotide sizes could reflect changes in preferred positions of the endonucleolytic cleavage that releases the SPO11-oligonucleotide complex, effects on 3'→5' exonucleolytic digestion of SPO11-oligonucleotides after they are formed, or occurrence of SPO11-induced DSBs at adjacent positions on the same DNA duplex (M. Neale, personal communication). Resection defects and

adjacent DSBs (which conventional cytology would be unable to resolve) are both possible explanations for why SPO11-oligonucleotide complexes in *Atm*^{-/-} spermatocytes show a greater increase than RAD51 focus numbers¹⁴.

Our results reveal an essential but previously unsuspected function for ATM in controlling the number of SPO11-generated DSBs. We suggest that activation of ATM by DSBs triggers a negative feedback loop that leads to inhibition of further DSB formation (Fig. 4b) via phosphorylation of SPO11 or its accessory proteins, several of which are known to be phosphorylated in budding yeast (for example, ref. 24) and are conserved in mammals². ATM is activated in the vicinity of DSBs, as judged by SPO11- and ATM-dependent appearance of γ H2AX (phosphorylated histone variant H2AX) on chromosomes at leptotene^{12,13,25}. Thus, we envision that the negative feedback loop operates at least in part at a local level, perhaps discouraging additional DSBs from forming close to where a DSB has already formed. Such a mechanism could minimize instances where both sister chromatids are cut in the same region, and could also promote more even spacing of DSBs along chromosomes. These studies provide a new molecular framework for understanding the gonadal phenotypes of patients with ataxia telangiectasia²⁶, which is caused by ATM deficiency²⁷.

METHODS SUMMARY

Mouse mutant alleles and the *Spo11* β transgene were previously described^{10,18,28–30}. Experimental animals were compared with controls from the same litter. Experiments conformed to regulatory standards and were approved by the MSKCC Institutional Animal Care and Use Committee. For measurement of SPO11-oligonucleotide complexes, both testes from each mouse were used per experiment, that is, littermate comparisons were made on a per-testis basis (Supplementary Fig. 1). Testis extract preparation, immunoprecipitation and western blot analysis were performed essentially as described⁷. Radiolabelled species were quantified with Fuji phosphor screens and ImageGauge software. The anti-mouse SPO11 monoclonal antibody was produced from hybridoma cell line 180 (M.P.T., unpublished data). The size distribution of SPO11-oligonucleotides was determined essentially as described⁵ after radiolabelling with [α -³²P] cordycepin. Benzonase treatment of SPO11-oligonucleotide complexes followed manufacturer's instructions (Novagen).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 June; accepted 25 August 2011.

Published online 16 October 2011.

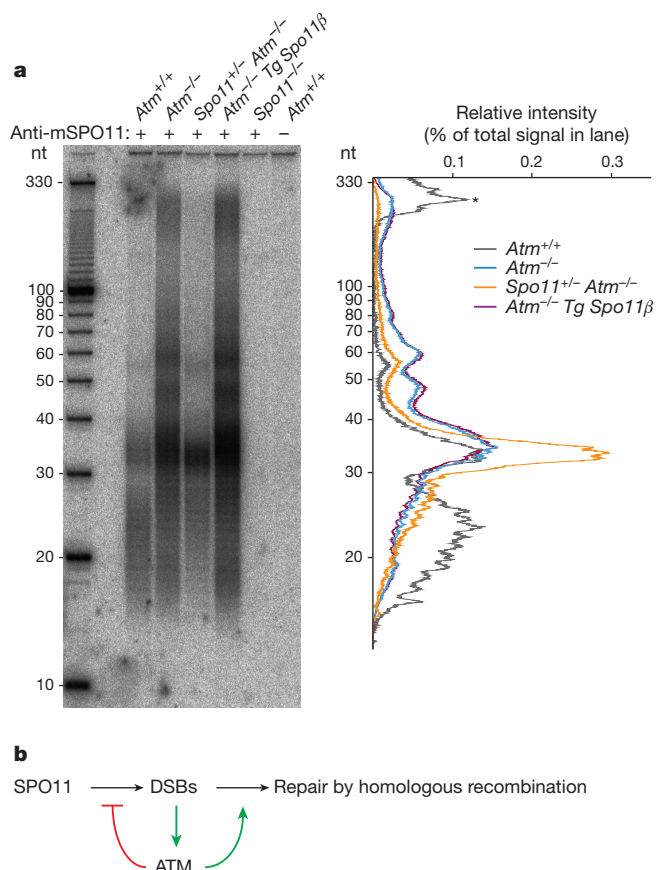


Figure 4 | Roles of ATM in DSB formation and processing. **a**, SPO11-oligonucleotide length distribution is altered in *Atm*^{-/-} spermatocytes. End-labelled SPO11-oligonucleotide complexes were treated with protease to digest the bound protein before electrophoresis on denaturing PAGE. Left, autoradiograph. Right, background-subtracted lane traces normalized to total signal within each lane. Asterisk, autoradiograph background. Each lane contains SPO11-oligonucleotides from the equivalent of different numbers of mice in order to better compare sizes: *Atm*^{+/-}, 15 mice; *Atm*^{-/-}, 2 mice; *Spo11*^{+/-} *Atm*^{-/-}, 4 mice; *Atm*^{-/-} plus transgene, 2 mice; *Spo11*^{-/-}, 2 mice; mock, 15 wild-type mice. nt, nucleotides. **b**, Negative feedback loop by which ATM regulates meiotic DSB levels. DSBs generated by SPO11 activate the ATM kinase, inhibiting further DSB formation. ATM may also have roles in repair of DSBs by homologous recombination; for example, by promoting DSB end resection.

- Keeney, S. in *Recombination and Meiosis: Crossing-Over and Disjunction* Vol. 2 (ed. Lankenau, D. H.) 81–123 (Springer, 2007).
- Cole, F., Keeney, S. & Jasin, M. Evolutionary conservation of meiotic DSB proteins: more than just Spo11. *Genes Dev.* **24**, 1201–1207 (2010).
- Sasaki, M., Lange, J. & Keeney, S. Genome destabilization by homologous recombination in the germ line. *Nature Rev. Mol. Cell Biol.* **11**, 182–195 (2010).
- Hochwagen, A. & Amon, A. Checking your breaks: surveillance mechanisms of meiotic recombination. *Curr. Biol.* **16**, R217–R228 (2006).
- Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
- Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).
- Daniel, K. *et al.* Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1. *Nature Cell Biol.* **13**, 599–610 (2011).
- Barlow, C. *et al.* *Atm*-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
- Xu, Y. *et al.* Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. *Genes Dev.* **10**, 2411–2422 (1996).
- Barlow, C. *et al.* *Atm* deficiency results in severe meiotic disruption as early as leptotene of prophase I. *Development* **125**, 4007–4017 (1998).
- Di Giacomo, M. *et al.* Distinct DNA-damage-dependent and -independent responses drive the loss of oocytes in recombination-defective mouse mutants. *Proc. Natl Acad. Sci. USA* **102**, 737–742 (2005).
- Barchi, M. *et al.* Surveillance of different recombination defects in mouse spermatocytes yields distinct responses despite elimination at an identical developmental stage. *Mol. Cell Biol.* **25**, 7203–7215 (2005).
- Bellani, M. A., Romanienko, P. J., Cairatti, D. A. & Camerini-Otero, R. D. SPO11 is required for sex-body formation, and Spo11 heterozygosity rescues the prophase arrest of *Atm*^{-/-} spermatocytes. *J. Cell Sci.* **118**, 3233–3245 (2005).

14. Barchi, M. *et al.* ATM promotes the obligate XY crossover and both crossover control and chromosome axis integrity on autosomes. *PLoS Genet.* **4**, e1000076 (2008).
15. Keeney, S. *et al.* A mouse homolog of the *Saccharomyces cerevisiae* meiotic recombination DNA transesterase Spo11p. *Genomics* **61**, 170–182 (1999).
16. Shannon, M., Richardson, L., Christian, A., Handel, M. A. & Thelen, M. P. Differential gene expression of mammalian *SPO11/TOP6A* homologs during meiosis. *FEBS Lett.* **462**, 329–334 (1999).
17. Bellani, M. A., Boateng, K. A., McLeod, D. & Camerini-Otero, R. D. The expression profile of the major mouse SPO11 isoforms indicates that SPO11 β introduces double strand breaks and suggests that SPO11 α has an additional role in prophase in both spermatocytes and oocytes. *Mol. Cell. Biol.* **30**, 4391–4403 (2010).
18. Kauppi, L. *et al.* Distinct properties of the XY pseudoautosomal region crucial for male meiosis. *Science* **331**, 916–920 (2011).
19. Romanienko, P. J. & Camerini-Otero, R. D. Cloning, characterization, and localization of mouse and human *SPO11*. *Genomics* **61**, 156–169 (1999).
20. Romanienko, P. J. & Camerini-Otero, R. D. The mouse *Spo11* gene is required for meiotic chromosome synapsis. *Mol. Cell* **6**, 975–987 (2000).
21. Bellve, A. R. *et al.* Spermatogenic cells of the prepubertal mouse. Isolation and morphological characterization. *J. Cell Biol.* **74**, 68–85 (1977).
22. Eaker, S., Cobb, J., Pyle, A. & Handel, M. A. Meiotic prophase abnormalities and metaphase cell death in MLH1-deficient mouse spermatocytes: insights into regulation of spermatogenic progress. *Dev. Biol.* **249**, 85–95 (2002).
23. Terasawa, M., Ogawa, T., Tsukamoto, Y. & Ogawa, H. Sae2p phosphorylation is crucial for cooperation with Mre11p for resection of DNA double-strand break ends during meiotic recombination in *Saccharomyces cerevisiae*. *Genes Genet. Syst.* **83**, 209–217 (2008).
24. Sasanuma, H. *et al.* Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. *Genes Dev.* **22**, 398–410 (2008).
25. Mahadevaiah, S. K. *et al.* Recombinational DNA double-strand breaks in mice precede synapsis. *Nature Genet.* **27**, 271–276 (2001).
26. Sedgwick, R. P. & Boder, E. in *Handbook of Clinical Neurology* Vol. 16 (ed. de Jong, J. M. B. V.) 347–423 (Elsevier, 1991).
27. Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
28. Edelmann, W. *et al.* Meiotic pachytene arrest in MLH1-deficient mice. *Cell* **85**, 1125–1134 (1996).
29. Pittman, D. L. *et al.* Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for *Dmc1*, a germline-specific RecA homolog. *Mol. Cell* **1**, 697–705 (1998).
30. Baudat, F., Manova, K., Yuen, J. P., Jasin, M. & Keeney, S. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol. Cell* **6**, 989–998 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Neale for discussions, R. Cha and K. McKim for sharing data before publication, and M. Hwang for assistance in monoclonal antibody development. This work was supported by NIH grants HD040916 and HD053855 (to M.J. and S.K.) and GM058673 (to S.K.). J.P. was supported in part by a Leukemia and Lymphoma Society Fellowship and F.C. by a Ruth L. Kirschstein NRSA (F32 HD51392). S.K. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions J.L., J.P. and F.C. performed experiments. M.P.T. generated the anti-SPO11 monoclonal hybridoma line. J.L., M.J., and S.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K. (s-keeney@ski.mskcc.org) or M.J. (m-jasin@ski.mskcc.org).

METHODS

Mouse mutant alleles and the *Spo11* β transgene were previously described^{10,18,28–30}. Experiments conformed to regulatory standards and were approved by the MSKCC Institutional Animal Care and Use Committee. For measurement of SPO11–oligonucleotide complexes, both testes from each mouse were used per experiment, that is, littermate comparisons were made on a per-testis basis (Supplementary Fig. 1). The anti-mouse SPO11 monoclonal antibody was produced from hybridoma cell line 180 (M.P.T., unpublished data).

Testis extract preparation, immunoprecipitation and western blot analysis were performed essentially as described⁷. Testes were decapsulated, then lysed in 800 μ l lysis buffer (1% Triton X-100, 400 mM NaCl, 25 mM HEPES-NaOH at pH 7.4, 5 mM EDTA). Lysates were centrifuged at 100,000 r.p.m. (355,040g) for 25 min in a TLA100.2 rotor. Supernatants were incubated with anti-mouse SPO11 antibody 180 (5 μ g per pair of testes) at 4 °C for 1 h, followed by addition of 30–40 μ l protein-A–agarose beads (Roche) and incubation for another 3 h. Beads were washed three times with IP buffer (1% Triton X-100, 150 mM NaCl, 15 mM Tris-HCl at pH 8.0). Immunoprecipitates were eluted with Laemmli sample buffer

and diluted six- to sevenfold in IP buffer. Eluates were incubated with additional anti-mouse SPO11 antibody 180 at 4 °C for 1 h, followed by addition of 30–40 μ l protein-A–agarose beads and incubation at 4 °C overnight. Beads were washed three times with IP buffer and twice with buffer NEB4 (New England BioLabs). SPO11–oligonucleotide complexes were radiolabelled at 37 °C for 1 h using terminal deoxynucleotidyl transferase (Fermentas) and [α -³²P] dCTP. Beads were washed three times with IP buffer, boiled in Laemmli sample buffer and fractionated on 8% SDS–PAGE. Complexes were transferred to a PVDF membrane by semi-dry transfer (Bio-Rad). Radiolabelled species were detected and quantified with Fuji phosphor screens and ImageGauge software. For western blot analysis, membranes were probed with anti-mouse SPO11 antibody 180 (1:2,000 in PBS containing 0.1% Tween 20 and 5% non-fat dry milk), then horseradish-peroxidase-conjugated protein A (Abcam; 1:10,000 in PBS containing 0.1% Tween 20 and 5% non-fat dry milk), and detected using the ECL+ reagent (GE Healthcare). The size distribution of SPO11–oligonucleotides was determined by radiolabelling with [α -³²P] cordycepin then protease digestion followed by denaturing PAGE. Benzoylase treatment of SPO11–oligonucleotide complexes was performed as per manufacturer's instructions (Novagen).

Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1

Valerie Garcia¹, Sarah E. L. Phelps¹, Stephen Gray¹ & Matthew J. Neale¹

Repair of DNA double-strand breaks (DSBs) by homologous recombination requires resection of 5'-termini to generate 3'-single-strand DNA tails¹. Key components of this reaction are exonuclease 1 and the bifunctional endo/exonuclease, Mre11 (refs 2–4). Mre11 endonuclease activity is critical when DSB termini are blocked by bound protein—such as by the DNA end-joining complex⁵, topoisomerases⁶ or the meiotic transesterase Spo11 (refs 7–13)—but a specific function for the Mre11 3'–5' exonuclease activity has remained elusive. Here we use *Saccharomyces cerevisiae* to reveal a role for the Mre11 exonuclease during the resection of Spo11-linked 5'-DNA termini *in vivo*. We show that the residual resection observed in Exo1-mutant cells is dependent on Mre11, and that both exonuclease activities are required for efficient DSB repair. Previous work has indicated that resection traverses unidirectionally¹. Using a combination of physical assays for 5'-end processing, our results indicate an alternative mechanism involving bidirectional resection. First, Mre11 nicks the strand to be resected up to 300 nucleotides from the 5'-terminus of the DSB—much further away than previously assumed. Second, this nick enables resection in a bidirectional manner, using Exo1 in the 5'–3' direction away from the DSB, and Mre11 in the 3'–5' direction towards the DSB end. Mre11 exonuclease activity also confers resistance to DNA damage in cycling cells, suggesting that Mre11-catalysed resection may be a general feature of various DNA repair pathways.

We sought to clarify Mre11-dependent processing reactions at DSBs with blocked termini. Blocked DSB ends are created in meiosis by the topoisomerase-like transesterase Spo11 to initiate crossover recombination¹⁴ (Supplementary Fig. 1a). Aside from the importance of meiotic recombination to sexual reproduction and genetic diversity, we reasoned that the molecular reactions pertaining to repair of covalently blocked Spo11–DSBs would be informative to the repair of DNA lesions with other types of end blockage, such as failed topoisomerase reactions and DSB ends competitively bound by the non-homologous end-joining (NHEJ) machinery.

Mre11-dependent endonucleolytic processing (nicking) of Spo11–DSBs generates two classes of Spo11–oligonucleotide fragments originating from the DSB end¹⁵ (Supplementary Fig. 1). Spo11–oligonucleotide complexes do not form in *mre11* mutants completely abrogated for nuclease activity (*mre11-D56N* and *mre11-H125N*; Fig. 1a), corroborating earlier conclusions that Mre11 directly processes Spo11–DSBs^{7–13}. Recently, a mutant of Mre11 that is exonuclease deficient, but mostly endonuclease proficient was described for the orthologous proteins in *Schizosaccharomyces pombe* and *Pyrococcus furiosus*¹⁶. We generated the equivalent mutation in *S. cerevisiae* Mre11 (His 59 to Ser; Supplementary Fig. 2a) and tested its function. During meiosis, Spo11–oligonucleotide complexes were readily detected in *mre11-H59S* cells (Fig. 1a), indicating that the allele is endonuclease proficient *in vivo*. Biochemical assays with recombinant Mre11-H59S showed reduced, but not abolished, 3'–5' exonuclease activity on linear duplex DNA, and retention of much of the single-stranded (ss)DNA endonuclease activity (Supplementary Fig. 2)—observations consistent with *mre11-H59S* partially separating the two nuclease functions.

We assessed DSB repair kinetics at two meiotic recombination hot-spots using Southern blotting and probes for the relevant genomic loci (*HIS4::LEU2* and *ARE1*; Fig. 1b and Supplementary Fig. 3). In the *mre11-H59S* strain, DSBs formed at normal levels and repaired as crossovers with normal timing (Fig. 1b, c). ssDNA resection, which can be qualitatively assessed by the relative migration of the DSB band on native agarose gels, also seemed unaffected by the *mre11-H59S* mutation (Fig. 1d, lanes 1 and 6, and Supplementary Fig. 4).

The lack of an obvious defect in ssDNA resection proficiency suggested that any potential contribution to ssDNA generation by Mre11 might be masked by the activity of another nuclease. During meiosis, the major resection pathway requires Exo1 (refs 17–21). We tested this idea by combining the *mre11-H59S* allele with an *EXO1* deletion (*exo1Δ*), which we found itself to have slightly delayed DSB repair kinetics, with fewer DSBs repairing as crossovers (Fig. 1b, c). Migration of the DSB band on agarose gels revealed that ssDNA resection was measurably reduced in *exo1Δ*—but not entirely abolished relative to a *mre11-H125N* control (where the failure to remove Spo11 prevents all resection; Fig. 1d, lanes 3, 4 and 5, and Supplementary Fig. 4). Our data agree with those of others investigating a meiotic role for Exo1 (refs 17–21).

The combination of *mre11-H59S* with *exo1Δ* caused DSBs to transiently accumulate for a longer period, with DSBs detectable for 2 h longer than in matched controls (Fig. 1b), and with formation of crossover recombinants reduced and delayed (Fig. 1c). These defects in DSB repair are correlated with a reduction in the mobility of DSB DNA on agarose gels (Fig. 1d and Supplementary Fig. 4). Specifically, in comparison to the *mre11-H59S* or *exo1Δ* single mutants, we observed DSB signals to migrate similarly to DSBs from the *mre11-H125N* control (Fig. 1d, lanes 2 and 3). This reduction in mobility is indicative of less ssDNA resection in *mre11-H59S/exo1Δ* than of either single mutant. We conclude that Exo1 and Mre11 collaborate to enable efficient ssDNA generation at meiotic DSBs.

Unrepaired DSBs activate Tel1 and Mec1 (yeast orthologues of the mammalian checkpoint kinases ATM and ATR respectively) to induce phosphorylation of histone H2Ax on Ser 129 and hyper-phosphorylation of Hop1 (a meiosis-specific adaptor of the DNA damage response²²). Phosphorylated H2Ax and Hop1 were detected in all strains, indicating that significant resection is not essential for checkpoint activation (Fig. 1e). However, phospho-H2Ax accumulated and persisted until late time points only in *mre11-H59S/exo1Δ*, and Hop1 phosphorylation persisted for at least 2 h longer than in matched controls. These observations are consistent with a genome-wide defect in DSB repair in *mre11-H59S/exo1Δ*. The Tel1 branch of the signalling pathway is primarily activated by unresected DSBs²³. We observed that Hop1 phosphorylation in the *mre11-H59S/exo1Δ* background is highly dependent on Tel1 (Supplementary Fig. 5), consistent with less resection occurring genome-wide.

To investigate if DSB processing defects were affecting meiotic chromosome segregation, we determined the efficiency of progression through anaphase I and II (Fig. 1f). Meiotic progression of wild type and *mre11-H59S* was essentially identical, whereas *exo1Δ* was delayed

¹Genome Damage and Stability Centre, The University of Sussex, Brighton, BN1 9RQ, UK.

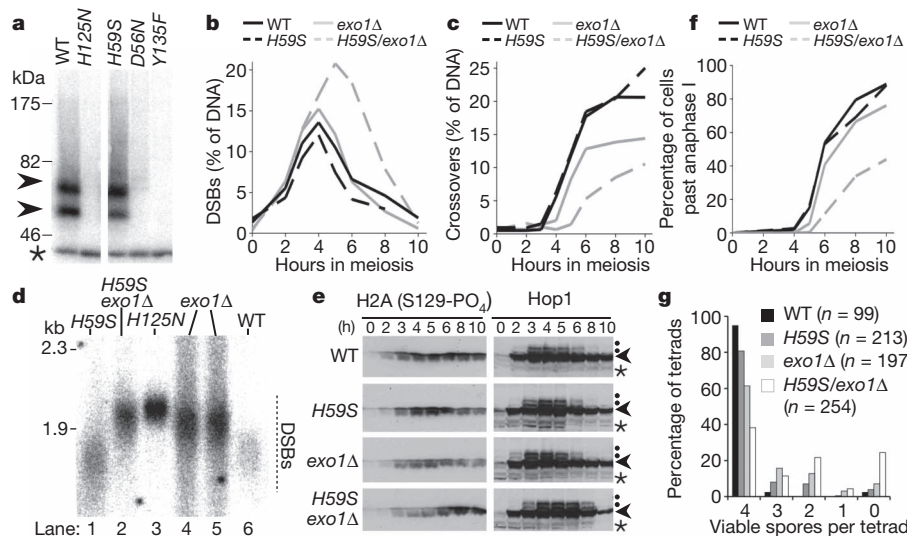


Figure 1 | Mre11- and Exo1-dependent resection and repair of meiotic DSBs. **a**, Autoradiograph of Spo11-oligonucleotide formation (arrowheads) in Mre11 nuclease-defective cells: *mre11-D56N*, *-H125N*, *-H59S*; and control *spo11-Y135F* cells in which Spo11-DSBs do not form. WT, wild type. Immunoprecipitated Spo11-oligonucleotide complexes are 3'-end-labelled using terminal transferase and separated on SDS-PAGE¹⁵ (see Supplementary Fig. 1). Asterisk marks an unrelated labelling artefact. **b–f**, Time course of events during meiosis for the indicated genotypes. **b**, **c**, Quantitative analysis of DSB (**b**) and crossover (**c**) signals at *HIS4:LEU2*. DNA at each time point was digested with PstI (**b**) or XhoI (**c**), separated by agarose gel electrophoresis and blots hybridized with probes to *LSB5* (**b**) or *STE50* (**c**). DSB signals were quantified as a percentage of specific lane signal (see also Supplementary Fig. 3).

by about 1 h, with slightly reduced overall efficiency. In contrast, nuclear division in the *mre11-H59S/exo1Δ* double mutant was poor, with more than half of the cells having failed to complete even the first meiotic nuclear division after 10 h in meiosis. Analysis of sporulation efficiency after 24 h revealed increased incidence of aberrant tetrad maturation and/or nuclear packaging where orphaned chromosome fragments were observed outside the maturing spore wall

d, Relative extent of DSB resection at *HIS4:LEU2* (4 h in meiosis). DNA was digested with BglII and blots hybridized with a *STE50* probe (see also Supplementary Fig. 4). **e**, Western blot analysis of phosphorylated H2A and Hop1 (arrowhead) detected from TCA-precipitated whole-cell lysates. Phosphorylated Hop1 is indicated by dot. Asterisk marks cross-reacting band. **f**, Progression through anaphase I and II assessed by microscopic examination of 4',6-diamidino-2-phenylindole (DAPI)-stained cells. **g**, Distribution of spore viabilities per tetrad (*n*, number of 4-spore tetrads dissected). The difference in distribution between *mre11-H59S/exo1Δ* and *exo1Δ* is highly significant (χ^2 test for goodness-of-fit; $\chi = 63.084$, 4 degrees of freedom, $P < 0.0001$). Absolute spore viabilities are: wild type, 97%; *mre11-H59S*, 90%; *exo1Δ*, 76%; *mre11-H59S/exo1Δ*, 59%.

(Supplementary Fig. 6). These defects in meiotic chromosome segregation manifested as reduced spore viability in the double mutant compared to controls (Fig. 1g).

To characterize in greater detail the molecular defect caused by *mre11-H59S*, we looked carefully at the distribution of Spo11-oligonucleotide products generated *in vivo* by the 5'-DSB processing reaction (Fig. 2a, b). In wild-type cells, two major classes of Spo11-oligonucleotide are observed, which differ by the length of attached DNA¹⁵. By contrast, in *mre11-H59S* we observed a shift in this distribution towards higher molecular weight Spo11-oligonucleotide species (Fig. 2a and Supplementary Fig. 7). Importantly, total Spo11-oligonucleotide formation was not itself delayed (Supplementary Fig. 7), indicating that the Spo11-removal reaction initiates with normal timing in *mre11-H59S* cells, but is defective in forming shorter molecules. To clarify the precise size distribution of the Spo11-oligonucleotide molecules, we fractionated de-proteinized oligonucleotides using denaturing polyacrylamide gel electrophoresis (PAGE; Fig. 2b). In wild-type cells, two peak areas of signal 10–17 nucleotides (nt) and 28–40 nt were apparent. These correspond to the shorter and longer Spo11-oligonucleotide classes detected on SDS-PAGE (Fig. 2a). We additionally detected signal (24% of total)

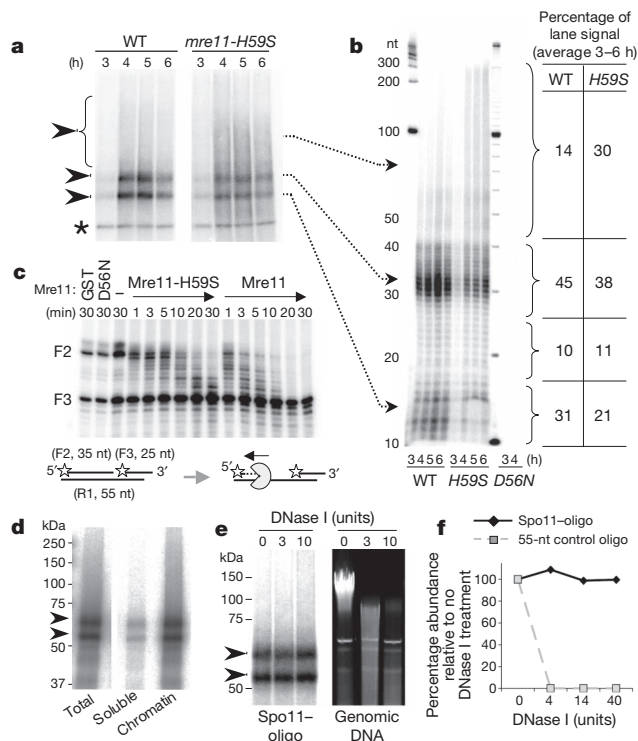


Figure 2 | Mre11-exonucleolytic processing of DSB ends. **a**, **b**, Spo11-oligonucleotide detection in wild type (WT) and *mre11-H59S* during meiosis. 3'-end-labelled Spo11 complexes are fractionated by SDS-PAGE (**a**) or by nucleotide resolution urea/PAGE after proteolytic removal of Spo11 peptide (**b**). **c**, *Mre11-H59S* shows reduced 3'-5' exonuclease activity on a nicked duplex. Reactions were performed as for Supplementary Fig. 2c. Stars indicate 5' label, F3 3' end has 5-nt extension and is refractory to Mre11-mediated resection²⁵. **d**, Chromatin association of Spo11-oligonucleotide complexes. Wild-type cell extracts were fractionated and the abundance of Spo11-oligonucleotide complexes assessed in soluble versus chromatin-enriched material. **e**, **f**, Nuclease resistance of Spo11-oligonucleotide (Spo11-oligo) complexes. Chromatin-enriched material from **d** was treated with DNase I, and abundance of Spo11-oligonucleotide complexes compared to the simultaneous degradation of genomic DNA (**e**), or of a control 55-nt oligonucleotide (**f**).

in molecules 18–27 and 41–300 nt long (Fig. 2b), indicating that the length distribution of processed molecules is significantly more heterogeneous than previously assumed, and that nicks are made at up to 300 nt from the DSB end.

In the *mre11-H59S* mutant, the distribution was shifted such that the long oligonucleotide molecules of 41–300 nt made up a third of the material detected (Fig. 2b). Although it is possible that the altered distribution of oligonucleotide molecules is caused by a reduction in the endonuclease activity of Mre11-H59S, our physical and genetic observations in both wild type and *mre11-H59S* can be readily explained if resection begins at relatively distant nicks (up to 300 nt from the DSB) and traverses bidirectionally both away from (using Exo1) and towards (using Mre11) the DSB end²¹. Such a model is consistent with the opposing polarities of the Mre11 and Exo1 exonuclease activities^{24,25}, and with the synergistic loss in resection we observe in *mre11-H59S/exo1Δ*. Moreover, the length distribution of Spo11-oligonucleotides is compatible with the extent of Exo1-independent resection reported recently by others²¹ and, as predicted, Spo11-oligonucleotide length is unchanged by loss of *EXO1* (Supplementary Fig. 8). Lastly, only a low background of Spo11-oligonucleotide complexes are detected in endonuclease-defective *mre11-D56N*, ruling out the possibility that the long oligonucleotide molecules arise via an alternative nuclease nicking the 5' strand (Fig. 2b). To test this mechanism *in vitro*, we incubated Mre11 protein with a nicked duplex substrate designed to mimic this proposed *in vivo* reaction, and found that Mre11-H59S resected from the nick with lower efficiency than wild-type Mre11 (Fig. 2c).

Together, these observations led us to consider that the steady-state length of Spo11-oligonucleotide complexes might arise via the relative processivity of the 3'–5' Mre11 exonuclease and the relative sensitivity to nucleolytic degradation of DNA close to the DSB end. Spo11-DSB formation requires at least ten factors¹⁴, suggesting that a large protein complex may reside at—and protect—the DSB end. If this model were correct, we expected Spo11-oligonucleotide complexes to be associated with chromatin and resistant to nucleolytic degradation. We tested this idea by incubating a chromatin-enriched nuclear pellet from meiotic yeast cells with DNase I, then assessed the quantity of Spo11-oligonucleotide complexes remaining, relative to both bulk DNA and to an exogenous protein-free oligonucleotide included in parallel reactions (Fig. 2d–f). Greater than 90% of Spo11-oligonucleotides are found in the chromatin fraction and, remarkably, no loss in signal was observed despite extensive nucleolytic degradation of both the chromosomal DNA and the control oligonucleotide. We conclude that Spo11-oligonucleotide complexes are occluded from degradation even by exogenous nucleases—a prediction of our model.

In cycling cells, Mre11 nuclease activity promotes the onset of resection—a requisite for repair of DSBs by homologous recombination¹. To investigate a specific role for the Mre11 3'–5' exonuclease during DNA repair in cycling cells, we challenged yeast cells with exposure to DNA damaging agents. Similar to complete abrogation of the endo/exonuclease activities (*mre11-H125N*), reduced Mre11 exonuclease activity (*mre11-H59S*) sensitized cells to the DNA alkylating agent methyl methanesulphonate (MMS) and to the topoisomerase poison camptothecin (CPT; Fig. 3). Compared to an *MRE11* deletion,

however, *mre11-H59S* and *mre11-H125N* are themselves far less sensitive, consistent with physical interactions within the Mre11 complex being retained (Supplementary Fig. 9). In agreement with Mre11 endonuclease activity being unaffected in *mre11-H59S*, and allowing redundant processing pathways, combining *mre11-H59S* with a deletion of *EXO1* did not further sensitize cells to MMS (Supplementary Fig. 10). Together these observations indicate that the exonuclease activity of Mre11 is involved in the repair of various classes of DNA lesion.

Understanding the regulation of DSB repair is a complex issue involving multiple factors with overlapping roles. Here, we propose a biological function for the 3'–5' exonuclease activity of the evolutionarily conserved Mre11 protein. Previous work has indicated DSB resection to traverse unidirectionally¹. We propose a refined model that involves the coordination of two resection activities of opposing polarity: Exo1 away from the DSB and Mre11 towards the DSB end (Fig. 4). We favour the view that this exonuclease reaction begins at nicks created by the Mre11 endonuclease (in conjunction with Sae2) and which are positioned at variable distance from the DSB end, perhaps due to locus-specific chromatin architecture²⁶. Although our assay detects only the site of incision closest to the DSB end, Mre11 and Sae2 may create multiple nicks on the resecting strand^{17,21,27} (A. S. H. Goldman, personal communication) that, in combination with exonucleolytic processing, might further enhance resection efficiency. Lastly, the recent observation that the length and abundance of Spo11-oligonucleotide complexes is increased in *Atm*^{−/−} mice²⁸ (S. Keeney and M. Jasin, personal communication), suggests that Mre11 exonuclease activity may be an evolutionarily conserved feature directly regulated by ATM.

During meiosis, bidirectional processing may help to reinforce subsequent steps of repair, which at least in some cases seem to occur differentially on either side of the DSB^{26,29,30}. Our observation that liberated Spo11-oligonucleotide complexes remain chromatin bound and relatively protected provides a clue to potential mechanisms of end differentiation. For example, retention of proteins on one or both of the DSB ends could influence subsequent steps of repair¹⁵ (Fig. 4b). Lastly, our observation that Mre11-dependent incision occurs at some

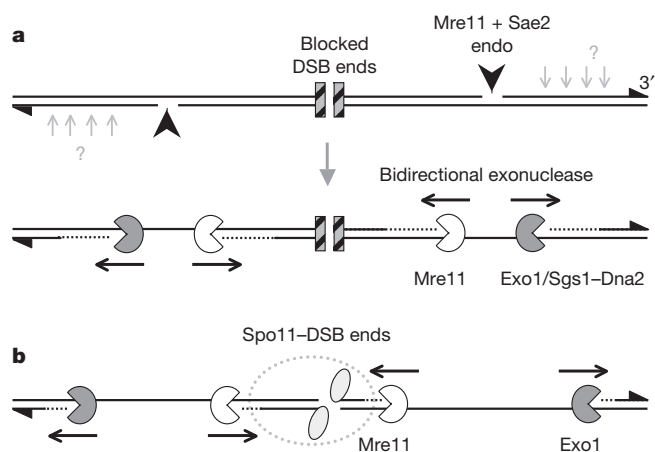


Figure 4 | Model for bidirectional processing of DSBs by Mre11 and Exo1.

a, After DSB formation with blocked ends (hatched squares), Mre11/Sae2-dependent nicks flanking the DSB ends create initiation sites for bidirectional resection by Exo1 and/or Sgs1–Dna2 away from the DSB, and by Mre11 towards the DSB end. Such terminal blocks could arise after base damage, trapping of a topoisomerase, or by avid binding of the NHEJ complex. 3' ends are marked with triangles. Mre11/Sae2 may make multiple nicks on the 5' strand (light grey arrows), facilitating resection. **b**, In meiosis, the DSB ends are terminally blocked by covalently bound Spo11 protein (grey ellipses), and may be protected from Mre11-dependent exonuclease degradation by a large metastable multisubunit complex (dashed outline), thereby generating the observed size distribution of Spo11-oligonucleotide complexes.

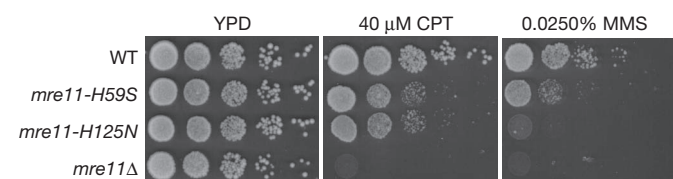


Figure 3 | DNA damage sensitivity of exonuclease-defective Mre11 cells. Tenfold serial dilutions of the indicated strains were spotted onto solid media containing the indicated compounds and incubated at 30 °C for 2 days (CPT) or 3 days (MMS).

distance from the DSB end indicates that DSB formation and processing reactions are coordinated over a considerable distance (300 bp of B-form DNA is ~100 nm). How these incision points are regulated—and restricted to only the 5'-ending strand—are fascinating questions for the future.

Although much of this work concerns specifics of meiotic DSB processing, we suggest that a similar pathway may occur whenever the DSB end is blocked by a lesion or protein complex that prevents direct loading of the 5'-3' exonuclease machinery. Such DNA blockages may be crosslinked protein, damaged DNA ends, or simply the stable binding of high affinity proteins to the DSB. As such, the nucleolytic incision pathway may provide the key point of regulation that controls the balance between NHEJ and homologous recombination.

METHODS SUMMARY

Yeast strains and culture methods. Meiotic cultures were prepared as described¹⁵. Strains were derived from SK1 using standard techniques. Spo11 protein is tagged by the HA3His6 epitope¹⁵. Mre11 mutations were introduced by pop-in/out using plasmids derived from pRS414-MRE11 (ref. 10). *exo1Δ*, *tel1Δ* and *sae2Δ* are full replacements of the open reading frame (ORF) with kanMX4, hphNT2 and kanMX6, respectively. A full strain list is available on request. For chronic DNA damage sensitivity, exponentially growing cultures were spotted in tenfold dilutions onto freshly prepared media.

Molecular techniques. Spo11-oligonucleotide complexes were detected by two rounds of immunoprecipitation and end labelling as described¹⁵. DSBs and cross-over recombinants were detected by Southern blotting genomic DNA after fractionation on agarose gels using standard techniques²⁹. Radioactive signals were collected on phosphor screens, scanned with a Fuji FLA5100, and quantified using ImageGauge software. Phosphorylated H2Ax and Hop1 protein were detected using Ab17353 (Abcam) and anti-Hop1 antisera (F. Klein), respectively. Subcellular fractionation of cell contents was performed by purifying hypotonically lysed spheroplasts through a sucrose cushion. Chromatin digests were performed at 25 °C with DNase I (NEB). Reactions were split and processed for DNA extraction or Spo11-oligonucleotide purification. As a positive control, parallel reactions included a 5'-labelled 55-nt oligonucleotide. Recombinant GST-Mre11 proteins were purified and reacted with DNA substrates as described¹⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 February; accepted 30 August 2011.

Published online 16 October 2011.

1. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349 (1999).
2. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770–774 (2008).
3. Nicolette, M. L. *et al.* Mre11–Rad50–Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks. *Nature Struct. Mol. Biol.* **17**, 1478–1485 (2010).
4. Zhu, Z., Chung, W. H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981–994 (2008).
5. Mimitou, E. P. & Symington, L. S. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J.* **29**, 3358–3369 (2010).
6. Hartsuiker, E., Neale, M. J. & Carr, A. M. Distinct requirements for the Rad32(Mre11) nuclease and Ctp1(Chp) in the removal of covalently bound topoisomerase I and II from DNA. *Mol. Cell* **33**, 117–123 (2009).
7. Furuse, M. *et al.* Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J.* **17**, 6412–6425 (1998).
8. Hartsuiker, E. *et al.* Ctp1^{Chp} and Rad32^{Mre11} nuclease activity are required for Rec12^{Spo11} removal, but Rec12^{Spo11} removal is dispensable for other MRN-dependent meiotic functions. *Mol. Cell. Biol.* **29**, 1671–1681 (2009).

9. Millman, N., Higuchi, E. & Smith, G. R. Meiotic DNA double-strand break repair requires two nucleases, MRN and Ctp1, to produce a single size class of Rec12 (Spo11)-oligonucleotide complexes. *Mol. Cell. Biol.* **29**, 5998–6005 (2009).
10. Moreau, S., Ferguson, J. R. & Symington, L. S. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol. Cell. Biol.* **19**, 556–566 (1999).
11. Nairz, K. & Klein, F. *mre11S*—a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes Dev.* **11**, 2272–2290 (1997).
12. Rothenberg, M., Kohli, J. & Ludin, K. Ctp1 and the MRN-complex are required for endonucleolytic Rec12 removal with release of a single class of oligonucleotides in fission yeast. *PLoS Genet.* **5**, e1000722 (2009).
13. Tsubouchi, H. & Ogawa, H. A novel *mre11* mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. *Mol. Cell. Biol.* **18**, 260–268 (1998).
14. Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52**, 1–53 (2001).
15. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
16. Williams, R. S. *et al.* Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135**, 97–109 (2008).
17. Hodgson, A. *et al.* Mre11 and Exo1 contribute to the initiation and processivity of resection at meiotic double-strand breaks made independently of Spo11. *DNA Repair* **10**, 138–148 (2010).
18. Keelagher, R. E., Cotton, V. E., Goldman, A. S. & Borts, R. H. Separable roles for Exonuclease I in meiotic DNA double-strand break repair. *DNA Repair* **10**, 126–137 (2010).
19. Manfrini, N., Guerini, I., Citterio, A., Lucchini, G. & Longhese, M. P. Processing of meiotic DNA double-strand breaks requires cyclin-dependent kinase and multiple nucleases. *J. Biol. Chem.* **285**, 11628–11637 (2010).
20. Tsubouchi, H. & Ogawa, H. Exo1 roles for repair of DNA double-strand breaks and meiotic crossing over in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **11**, 2221–2233 (2000).
21. Zakharyevich, K. *et al.* Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. *Mol. Cell* **40**, 1001–1015 (2010).
22. Carballo, J. A., Johnson, A. L., Sedgwick, S. G. & Cha, R. S. Phosphorylation of the axial element protein Hop1 by Mec1/Tel1 ensures meiotic interhomolog recombination. *Cell* **132**, 758–770 (2008).
23. Usui, T., Ogawa, H. & Petrini, J. H. A. DNA damage response pathway controlled by Tel1 and the Mre11 complex. *Mol. Cell* **7**, 1255–1266 (2001).
24. Szankasi, P. & Smith, G. R. A. DNA exonuclease induced during meiosis of *Schizosaccharomyces pombe*. *J. Biol. Chem.* **267**, 3014–3023 (1992).
25. Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Mol. Cell* **1**, 969–979 (1998).
26. Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).
27. Jazayeri, A., Balestrini, A., Garner, E., Haber, J. E. & Costanzo, V. Mre11–Rad50–Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. *EMBO J.* **27**, 1953–1962 (2008).
28. Lange, J. *et al.* ATM controls meiotic double-strand-break formation. *Nature* doi:10.1038/nature10508 (this issue).
29. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell* **106**, 59–70 (2001).
30. Kim, K. P. *et al.* Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. *Cell* **143**, 924–937 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Cha, E. Hoffmann, N. Hunter, S. Keeney and J. Nitiss for yeast strains; L. Symington for plasmids; F. Klein and J. Petrini for antisera. V.G. is supported by an MRC New Investigator Grant to M.J.N. M.J.N. is supported by a University Research Fellowship from the Royal Society and a Career Development Award from the Human Frontiers Science Program Organisation.

Author Contributions V.G. and M.J.N. designed the experiments and wrote the paper. V.G. and M.J.N. performed the experiments with technical support from S.E.L.P. and S.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.J.N. (m.neale@sussex.ac.uk).

METHODS

Yeast strains. Strains of *S. cerevisiae* were derived from SK1 using standard techniques. Spo11 protein is tagged by the HA3His6 epitope¹⁵. *Mre11* mutations were introduced by pop-in/out using plasmids derived from pRS414-MRE11 (see below). *exo1Δ*, *tel1Δ* and *sae2Δ* are full replacements of the open reading frame (ORF) with kanMX4, hphNT2 and kanMX6, respectively. A full strain list is available on request.

Meiotic cultures. YPD cultures (1% yeast extract, 2% peptone, 2% glucose) were diluted 100-fold into YPA (1% yeast extract, 2% peptone, 1% K-acetate) and grown vigorously for 14 h at 30 °C. Cells were collected by centrifugation, washed once in water, resuspended in an equal volume of prewarmed 2% K-acetate containing diluted amino acid supplements, and shaken vigorously at 30 °C.

Construction of *Mre11* nuclease mutant strains. The *mre11-H59S* mutant was constructed by site-directed mutagenesis PCR on the integrating plasmid pSM444 (pRS406::*mre11-D56N*), reversing the D56N mutation to wild type and introducing H59S (forward primer: GTACAGTCCGGTATCTTTTACGCTG AATAAGCC; reverse primer: GGCTTATTCACGCTAAAAAGATCACCGG ACTGTAC). Integration plasmids containing *mre11-D56N* and *mre11-H125N* nuclease dead mutant alleles were provided by L. Symington (pSM444 and pSM438 respectively). To replace chromosomal *MRE11* with mutant alleles, the plasmid containing mutations in the nuclease domains of *MRE11* were linearized with SphI and transformed in the SK1 strain. Uracil-positive (Ura⁺) transformants were inoculated in rich medium overnight and 20 µl of culture were spread onto medium containing 5-fluoroorotic acid (5-FOA) in order to select pop-out events. The presence of the *mre11-D56N* and *mre11-H125N* alleles in the resulting 5-FOA resistant cells was assessed by sensitivity to DNA damaging agent: cells from single 5-FOA resistant colonies were patched onto plates containing high concentration of CPT (30–50 µM) supplemented with phloxin B. Because the introduction of the *mre11-H59S* mutation also altered a PmlI restriction site, the presence of the *mre11-H59S* allele was tested by sensitivity of PCR reactions covering the mutated region to PmlI restriction. Replacement of wild-type *MRE11* by targeted mutations was confirmed by PCR amplification and sequencing.

Spot tests for DNA damage sensitivity. Cells were grown overnight in liquid YPD, diluted into fresh media and grown to log phase, adjusted to OD_{600 nm} = 0.2, serially diluted tenfold, and 5 µl spotted onto control and drug-containing YPD plates.

Spo11–oligonucleotide assays. Spo11–oligonucleotide complexes were detected by immunoprecipitation and end labelling following established methods¹⁵. To reduce the co-precipitation of nonspecific genomic DNA, two rounds of immunoprecipitation were used. Specifically, 10–50 ml of sporulating culture was lysed in 10% ice-cold TCA using zirconia beads and a BioSpec 24. Precipitated material was dissolved in STE (2% SDS, 0.5 M Tris pH 8.1, 10 mM EDTA, 0.05% bromophenol blue), and boiled for 5 min. Extracts were diluted twofold in 2× IP buffer (2% Triton X-100, 300 mM NaCl, 30 mM Tris-HCl pH 8.1, 2 mM EDTA), centrifuged for 10 min at 16,000g at 4 °C, and supernatant was diluted a further twofold in 1× IP buffer. Anti-HA antibody (F-7; Santa Cruz Biotechnology) was added at 1 in 500, protein-G-agarose matrix (Roche) at 1 in 50, and then incubated with rotation for 4 h at 4 °C. Immune complexes were collected by low speed centrifugation, and washed three times with 1× IP buffer. Beads were boiled for 5 min in 250 µl STE, chilled on ice, and diluted twofold as above with 2× IP buffer, recentrifuged, and supernatant was dissolved further twofold in 1× IP buffer. Fresh antibody and beads were added at the above dilutions. The second immunoprecipitation was performed overnight at 4 °C, and then washed as above. Two additional washes in 1× TKAC (20 mM Tris-acetate pH 7.9, 50 mM K-acetate) were performed before incubation with 10–20 units TdT (Fermentas) and 5–10 µCi CoTP (cordycepin triphosphate; Perkin Elmer) in 1× TKAC, 0.25 mM CoCl₂ buffer at 37 °C for 1 h. Labelled complexes were washed twice with 1× IP buffer, and eluted in Laemmli loading buffer for direct analysis on 7.5% SDS–PAGE. For nucleotide resolution analysis of Spo11–oligonucleotide lengths, eluted complexes were mixed with 1 µg glycogen (Roche) and 10 volumes of 100% ethanol, and precipitated overnight at –20 °C. Precipitates were collected by centrifugation, dissolved in 15 µl of 10× TE containing 0.5 µg ml^{–1} proteinase K, and incubated at 50 °C for 30 min. Eluted oligonucleotides were mixed with 3 volumes of loading dye (95% formamide, 10 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol), and fractionated through a 28-cm tall, 0.5-mm thick 12% polyacrylamide (19:1), 6 M urea gel in 1× TBE running buffer at approx 1,200 V for 50–60 min. Gels were fixed in 10% methanol, 7% acetic acid and 5% glycerol, vacuum dried and exposed to phosphor screens for imaging.

DSB and crossover analysis. Genomic DNA was isolated from aliquots of synchronously sporulating cultures using standard methods. Briefly, spheroplasts were prepared in 1 M sorbitol, 0.1 M EDTA, 0.1 M NaH₂PO₄ pH 7.5, 1% BME and 200 µg ml^{–1} zymolyase 100T for 20–30 min at 37 °C, and lysed by adding SDS to 0.5% and proteinase K to 200 µg ml^{–1} with incubation for 60 min at 60 °C. Protein

was removed by mixing with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), and nucleic acids precipitated by adding one-tenth volume of 3 M NaAc pH 5.2 and an equal volume of 100% ethanol. Precipitates were washed in 70% ethanol and dissolved in 1× TE containing 100 µg ml^{–1} RNase, incubated for 60 min at 37 °C, reprecipitated with ethanol and NaAc and DNA pellets left to dissolve in 1× TE overnight at 4 °C. Genomic DNA was digested with restriction enzymes following the manufacturer's recommendations, fractionated on 0.6%, 0.7% or 0.9% agarose gels in 1× TAE or 1× TBE buffer and ethidium bromide (as required), and blotted under vacuum to zeta probe membrane (BioRad) in 0.5 M NaOH, 1.5 M NaCl. Blots were equilibrated in hybridization solution (0.5 M NaH₂PO₄ pH 7.2, 7% SDS, 1 mM EDTA, 1% BSA) at 65 °C, and random primed radioactive probes (High Prime; Roche) prepared from gel-purified PCR products were then added. Blots were washed four times in 40 mM NaH₂PO₄, 1% SDS, 1 mM EDTA at 65 °C, then dried and exposed to phosphor screens for imaging.

Western blotting. TCA-denatured cell material dissolved in STE and 5% β-mercaptoethanol was fractionated in 7.5% or 10% SDS–PAGE, transferred to PVDF membrane (Millipore) in 1× CAPS buffer (10 mM CAPS–NaOH pH 11, 10% methanol), and incubated in TBST + 3% BSA with the requisite antibodies: phosphorylated H2Ax was detected using Ab17353 diluted 1/5,000 in TBST (Abcam) and Hop1 detected using anti-Hop1 antisera at a 1/4,000 dilution³¹.

DAPI staining. Cells were fixed in 100% methanol, and aliquots mixed with 1 µg ml^{–1} DAPI in 75% glycerol. The number of cells with 0, 2 and 4 nuclei was scored under a fluorescence microscope. At least 200 cells were counted per time point. Sporulation efficiency was determined by observation of fixed tetrads (24 h into meiosis, on three independent time courses) under a Zeiss fluorescence microscope with both fluorescence and bright-field illumination. Samples were randomized and 150 to 300 cells were counted for each strain per time course. Sporulations were scored as abnormal when significant extranuclear material was detected by DAPI staining and the wall morphology of spores was abnormal under bright-field illumination. DAPI and DIC images were captured using a personalDV (DeltaVision) system (Applied Precision) using the softWoRx software.

Chromatin enrichment. Cell pellets corresponding to 100 ml meiotic cultures were resuspended in 30 ml Sphero/CoHex buffer (1 M sorbitol, 50 mM HEPES pH 7.5, 10 mM EDTA, 5 mM hexamine cobalt chloride) supplemented with 1% β-mercaptoethanol and zymolyase 100T to 50 µg ml^{–1}, and incubated for ~25 min at 37 °C. Spheroplasts were transferred onto 10 ml sucrose cushion, spun for 5 min at 16,000g at 4 °C. Spheroplast pellets were resuspended in 1 ml sucrose buffer (1 M sucrose, 50 mM MES–NaOH, 50 mM NaCl, 1 mM EDTA, 0.5 mM MgCl₂, Roche protease inhibitors cocktail, AEBSEF to 10 µg ml^{–1}), lysed in 30 ml lysis buffer (50 mM MES–NaOH, 1 mM MgCl₂, 5 mM EDTA, Roche protease inhibitors cocktail, AEBSEF to 10 µg ml^{–1}), homogenized in a dounce homogenizer, and spun 5 min at 16,000g at 4 °C. Pellets were resuspended in 30 ml lysis buffer and subcellular fractionation of cell contents was performed by centrifugation through a 10-ml sucrose cushion. Pellets were washed once in the same buffer, resuspended in sucrose buffer, and 300 µl aliquots were stored at –20 °C. Aliquots were incubated at 25 °C with DNase I (NEB) in 50 mM HEPES, 50 mM NaCl, 5 mM MgCl₂ buffer supplemented with complete protease inhibitors tablet (Roche) and AEBSEF. Reactions were split and processed for genomic DNA extraction or Spo11–oligonucleotide purification. As a positive control, parallel reactions included a 5'-labelled 55-base oligonucleotide (R1) that was detected on 1× TBE PAGE.

Recombinant protein purification. GST–*Mre11* proteins were purified and reacted with DNA substrates as described¹⁰. Briefly, *MRE11* wild-type, *mre11-H59S* and *mre11-D56N* open reading frames (ORFs) were amplified by PCR, cloned with BamHI/Sall ends into pEG-KT plasmid and were transformed into the protease defective JEL1 *mre11Δ* strain (LSY1706; a gift from L. Symington). Galactose induction was performed as described in <http://www.bio.brandeis.edu/haberlab/jehsite/protocol.html>. Cells from a 100 ml culture were collected by centrifugation for 5 min at 3,000g at 4 °C, washed in cold lysis buffer (20 mM Tris-HCl pH 8, 1 mM EDTA, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol), and lysed in 500 µl lysis buffer supplemented with 50 mM NaF, 10 µg ml^{–1} AEBSEF and complete protease inhibitor tablet (Roche) using zirconia beads and a BioSpec 24 apparatus. Lysates were cleared by centrifugation for 10 min at 16,000g at 4 °C, and supernatants were incubated with 200 µl of Glutathione Sepharose beads slurry (pre-washed in lysis buffer) for 1 h at 4 °C. Beads were washed three times in lysis buffer and proteins were eluted over 30 min at 4 °C in elution buffer (50 mM Tris pH 8, 1 mM EDTA, 20 mM glutathione reduced, 40% glycerol, supplemented with protease inhibitors), and aliquots were frozen in liquid nitrogen and stored at –80 °C.

Oligonucleotide nicked substrate and exonuclease assay. The nicked, double-stranded substrate described in Fig. 2c was made by 5'-[³²P] end labelling of oligonucleotides F2: GACCTGGCAGCTAGGACAGCATGGGATCTGGCCTG; and F3: TGTTACACAGTGCTACAGACatggt (lowercase indicates 3' overhang) with PNK; and annealing to the unlabelled reverse complement, R1:

GTCTGTAGCACTGTGTAACACAGGCCAGATCCCATGCTGTCCTACGTGC CAGGT. Annealing was verified by migration on 1× TBE, 12% polyacrylamide gel. ~0.66 pmol of double-stranded nicked substrate was incubated with ~0.13 pmol of recombinant protein in nuclease buffer (25 mM MOPS pH 7, 60 mM KCl, 0.2% Triton X-100, 2 mM DTT, 50 mM MnCl₂) for the indicated length of time at 37 °C. Reactions were stopped by adding SDS to 0.3% and proteinase K to 0.5 mg ml⁻¹ and incubated for 15 min at 37 °C. Reactions were

denaturated in 2 volumes of formamide loading buffer (95% formamide, 10 mM EDTA, 0.005% bromophenol blue, 0.01% xylene cyanol) at 95 °C and separated on a 6 M urea, 12% polyacryamide gel in 1× TBE. Gels were exposed to a phosphor screen for imaging.

31. Panizza, S. *et al.* Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell* **146**, 372–383 (2011).

Rad51 paralogues Rad55–Rad57 balance the antirecombinase Srs2 in Rad51 filament formation

Jie Liu¹, Ludovic Renault², Xavier Veaute³, Francis Fabre³, Henning Stahlberg^{2,4} & Wolf-Dietrich Heyer^{1,2}

Homologous recombination is a high-fidelity DNA repair pathway. Besides a critical role in accurate chromosome segregation during meiosis, recombination functions in DNA repair and in the recovery of stalled or broken replication forks to ensure genomic stability. In contrast, inappropriate recombination contributes to genomic instability, leading to loss of heterozygosity, chromosome rearrangements and cell death. The RecA/UvsX/RadA/Rad51 family of proteins catalyses the signature reactions of recombination, homology search and DNA strand invasion^{1,2}. Eukaryotes also possess Rad51 paralogues, whose exact role in recombination remains to be defined³. Here we show that the *Saccharomyces cerevisiae* Rad51 paralogues, the Rad55–Rad57 heterodimer, counteract the antirecombination activity of the Srs2 helicase. The Rad55–Rad57 heterodimer associates with the Rad51–single-stranded DNA filament, rendering it more stable than a nucleoprotein filament containing Rad51 alone. The Rad51–Rad55–Rad57 co-filament resists disruption by the Srs2 antirecombinase by blocking Srs2 translocation, involving a direct protein interaction between Rad55–Rad57 and Srs2. Our results demonstrate an unexpected role of the Rad51 paralogues in stabilizing the Rad51 filament against a biologically important antagonist, the Srs2 antirecombination helicase. The biological significance of this mechanism is indicated by a complete suppression of the ionizing radiation sensitivity of *rad55* or *rad57* mutants by concomitant deletion of *SRS2*, as expected for biological antagonists. We propose that the Rad51 presynaptic filament is a meta-stable reversible intermediate, whose assembly and disassembly is governed by the balance between Rad55–Rad57 and Srs2, providing a key regulatory mechanism controlling the initiation of homologous recombination. These data provide a paradigm for the potential function of the human RAD51 paralogues, which are known to be involved in cancer predisposition and human disease.

Rad51 protein and its homologues RecA, UvsX and RadA form nucleoprotein filaments with ssDNA that perform homology search and DNA strand invasion during homologous recombination. The Rad51 paralogues share the RecA core with the Rad51 protein featuring unique amino- and carboxy-terminal extensions (Supplementary Fig. 2), but themselves do not form filaments and are unable to perform homology search and DNA strand invasion^{2–4}. Whereas humans contain five paralogues (RAD51B, RAD51C, RAD51D, XRCC2, XRCC3), the budding yeast *Saccharomyces cerevisiae* contains two clearly identifiable paralogues, Rad55 and Rad57 (Supplementary Fig. 2). Rad55 and Rad57 in yeast as well as the five human RAD51 paralogues have unique non-redundant functions in recombination, and mutations in any one of them lead to recombination defects, chromosomal instability, sensitivity to DNA damage, and meiotic defects^{1–3}. Defects in the budding yeast *RAD55* and *RAD57* genes lead to identical and epistatic phenotypes in DNA repair and recombination, consistent with the formation of a stable Rad55–Rad57 heterodimer^{4,5}. Rad55–Rad57 heterodimers were inferred to function as

mediator proteins⁶ allowing assembly of the Rad51 nucleoprotein filament on ssDNA covered by the eukaryotic ssDNA-binding protein RPA⁴. This suggested that Rad55–Rad57 are involved in the nucleation of the Rad51 filament, which is otherwise inhibited on RPA-covered ssDNA. This nucleation model is akin to the role of RecFOR or BRCA2 in nucleating RecA or human RAD51 filaments^{7–9}. Rad51 filament formation *in vivo* can be monitored cytologically as Rad51 focus formation at the site of DNA damage¹⁰. Unexpectedly, Rad51 focus formation after ionizing radiation in yeast was demonstrated to be independent of Rad55–Rad57 and formation of visible Rad55–Rad57 foci required Rad51 (ref. 10). These results are difficult to reconcile with the nucleation model derived from the biochemical results and suggest an alternative function of Rad55–Rad57 *in vivo*.

To address the function of the Rad51 paralogues in yeast, we determined the effect of Rad55–Rad57 on the stability of Rad51–ssDNA nucleoprotein complexes. Deletion mutants of the *RAD55* or *RAD57* genes display a curious enhancement of some phenotypes at low temperature (in particular ionizing radiation sensitivity; see Supplementary Fig. 12)⁵, indicating that these proteins are involved in the stabilization of a molecular complex, probably the Rad51 presynaptic filament. To test this hypothesis, we incubated subsaturating amounts of Rad51 protein with ssDNA (1 Rad51 per 15 nucleotides) in the presence of substoichiometric amounts of Rad55–Rad57 heterodimer (1 Rad55–Rad57 per 4 Rad51) and challenged the filaments with buffer containing a high salt concentration (500 mM NaCl) (Supplementary Fig. 3a, b). Under these conditions, Rad51 does not maintain stable complexes with ssDNA during electrophoresis. However, the presence of Rad55–Rad57 resulted in stable, Rad51-containing ssDNA complexes that withstood the salt challenge. In a complementary approach, we examined the effect of Rad55–Rad57 on Rad51 filament formation at near-physiological ionic strength (90 mM NaCl) (Fig. 1a, b). Under these conditions, only a fraction of the available Rad51 binds ssDNA, causing retarded mobility of the DNA (Fig. 1b, lane 3). Addition of substoichiometric amounts of Rad55–Rad57 (1 Rad55–Rad57 per 6 Rad51 in lane 4 of Fig. 1b) led to the formation of a novel, supershifted complex that contained both Rad51 and Rad55–Rad57, as demonstrated by immunoblotting. Rad55–Rad57 heterodimer alone binds to DNA under these conditions, leading to the formation of protein networks that are too large to enter the gel (Fig. 1b, lane 2). The results from both experiments (Fig. 1b; Supplementary Fig. 3) indicate that Rad55–Rad57 form a co-complex with Rad51 on ssDNA and stabilize Rad51–ssDNA filaments. Indeed, immunogold electron microscopy targeted towards Rad55 (glutathione S-transferase (GST)-tag; see Fig. 1c) directly visualized Rad55 associated with the Rad51–ssDNA filaments (Fig. 1d). Control experiments demonstrated the specificity of the gold labelling (Supplementary Table 1) with over 90% of the gold particles associated with clearly identifiable Rad51 filaments. The remainder may have associated with filaments too short to be scored or with free Rad55–Rad57. Gold particles were found either at the filament terminus ($n = 40$) or

¹Department of Microbiology, University of California, Davis, Davis, California 95616-8665, USA. ²Department of Molecular & Cellular Biology, University of California, Davis, Davis, California 95616-8665, USA. ³CEA-DSV-Institut de Radiobiologie Cellulaire et Moléculaire, UMR217 CNRS/CEA, F-92265 Fontenay aux Roses, France. ⁴Center for Cellular Imaging and Nanoanalytics, University Basel, CH-4056 Basel, Switzerland.

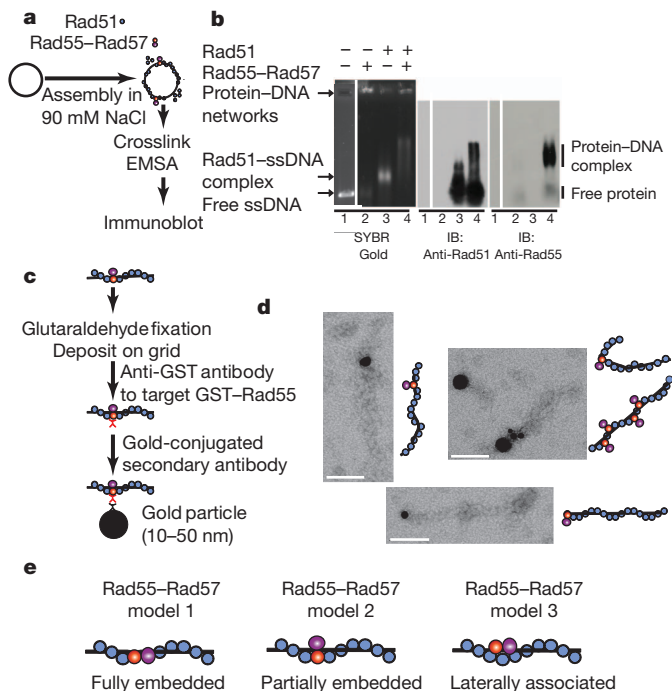


Figure 1 | Rad55–Rad57 is associated with and stabilizes Rad51–ssDNA filaments. **a**, Rad51–ssDNA filament assembly assay. EMSA, electrophoretic mobility shift assay. **b**, Rad51 (0.67 μM) with or without Rad55–Rad57 (0.11 μM) was incubated with 4 μM ϕX174 ssDNA. The migration position of free protein was confirmed in controls lacking DNA (Supplementary Fig. 3c). IB, immunoblot. **c**, Reaction scheme of immunoaffinity gold labelling of Rad55. **d**, Electron microscopy images of gold-labelled Rad55 associated with Rad51–ssDNA filament (1:3 Rad51:nucleotide; 2.34 μM Rad51 \pm 0.43 μM Rad55–Rad57, 7 μM ssDNA). Scale bars, 100 nm. **e**, Models for the disposition of Rad55–Rad57 with the Rad51 filament. For simplicity, only model 2 is drawn in further illustrations.

interstitially ($n = 43$) (Supplementary Table 1). Negative controls with Rad51 filaments assembled in the absence of Rad55–Rad57 showed negligible gold labelling (Supplementary Table 1). These data show that Rad55–Rad57 are associated with the Rad51–ssDNA filament, but the exact disposition of the heterodimer with the filament remains to be determined (see Fig. 1e).

Salt stability of protein–DNA complexes is a valuable biochemical criterion. To establish biological significance, we tested whether Rad55–Rad57 heterodimers stabilize Rad51–ssDNA filaments against a biologically relevant destabilizer. The Srs2 helicase was identified as a negative regulator of homologous recombination, and genetic experiments indicated that Srs2 targets Rad51 protein^{11–13}. Consistent with the genetic data, Srs2 translocates on ssDNA and disrupts Rad51 pre-synaptic filaments *in vitro*, providing a compelling mechanism for its function as an antirecombinase^{14–16}. In the presence of 0.1 or 0.33 μM Srs2 approximately 70% of the Rad51 is dissociated as assessed by measuring Rad51 associated with ssDNA coupled to magnetic beads (Fig. 2a–c). The presence of substoichiometric amounts of Rad55–Rad57 (0.1 μM) enhanced the recovery of ssDNA-bound Rad51 by ~twofold (from 31% to 60% in the presence of 0.33 μM Srs2). Rad55–Rad57 and Srs2 bound to Rad51-covered ssDNA in a quantitative and concomitant manner (Fig. 2d). Together the data show that Rad55–Rad57 inhibit Srs2 when bound to DNA and not in solution. Concentration-dependent inhibition of Srs2-mediated dissociation of Rad51 from ssDNA by Rad55–Rad57 was also observed in a topology-based assay (Supplementary Figs 4, 5).

To investigate the role of Rad55–Rad57 in antagonizing disruption of Rad51 presynaptic filaments by Srs2 further, we used electron microscopy to examine nucleoprotein filaments directly (Fig. 3 and

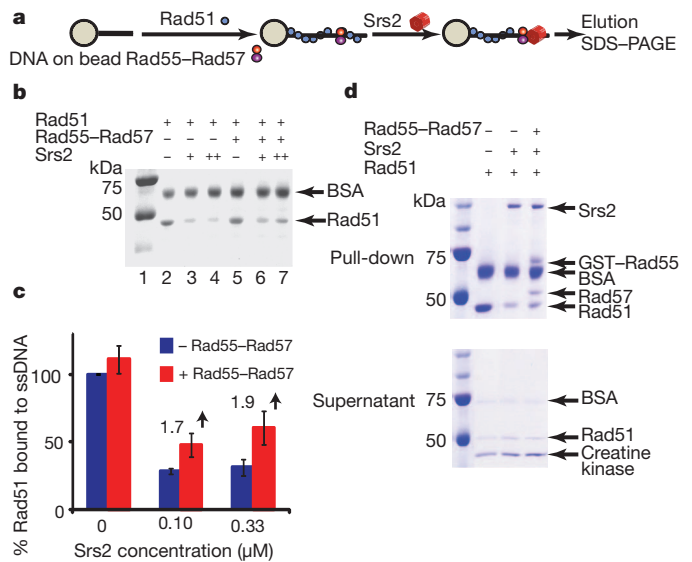


Figure 2 | Rad55–Rad57 stabilize Rad51–ssDNA filaments to resist disruption by Srs2. **a**, Pull-down assay measuring stability of Rad51–ssDNA complexes (1:3 Rad51:nucleotide, 1 μM Rad51 \pm 0.1 μM Rad55–Rad57) against disruption by Srs2 (0.1 or 0.33 μM). **b**, Rad51 remaining bound to ssDNA. **c**, Quantification of results in **b** and additional experiments. Shown are means \pm 1 s.d., $n = 3$. **d**, Concomitant binding of Rad55–Rad57 and Srs2 to Rad51-covered ssDNA. Pull-down assay measuring stability of Rad51–ssDNA complexes (1:3 Rad51:nucleotide, 1 μM Rad51 \pm 0.2 μM Rad55–Rad57) against disruption by 0.33 μM Srs2. Top, pull-downs; bottom, supernatants.

Supplementary Fig. 6). Rad51 filaments were assembled on a 600-nucleotide fragment of ssDNA and RPA was added to visualize free ssDNA. Consistent with previous observations^{14,15}, in the absence of Rad55–Rad57 Srs2 disrupts the Rad51–ssDNA filament efficiently, leading to binding of RPA to the newly exposed ssDNA (Fig. 3). Importantly, when substoichiometric amounts of Rad55–Rad57 were incubated with Rad51 and ssDNA, the filaments were stabilized against disruption by Srs2, as indicated by the significantly increased mean filament length.

How do Rad55–Rad57 heterodimers block Srs2 from dissociating Rad51 from ssDNA? Srs2 is known to interact with Rad51 and trigger the Rad51 ATPase leading to dissociation of Rad51 from ssDNA¹⁶. We found that Rad55–Rad57 form a 1:1 complex with Srs2 (Fig. 4a) and have higher affinity to Srs2 than to Rad51 (Fig. 4b, c). Excess Rad51 does not compete with Srs2 binding to Rad55–Rad57 (Supplementary Fig. 7). Moreover, Rad55–Rad57 heterodimers are able to simultaneously bind Rad51 and Srs2 in a 1:1:1 stoichiometry (Fig. 4d and Supplementary Figs 7–9). We considered the possibility that Rad55–Rad57 inhibit the Srs2 ATPase activity and by that Srs2 translocation, but Srs2 ATPase activity is barely altered by the presence of Rad55–Rad57 (data not shown). Srs2 translocase/helicase activity is stimulated by Rad51 binding to DNA¹⁷ (Fig. 4e–g). Importantly, Rad55–Rad57 completely suppress this stimulatory effect of Rad51, leading to inhibition of the Srs2 helicase activity even at a fivefold molar excess of Srs2 over Rad55–Rad57 (Fig. 4f, g and Supplementary Fig. 10). This substoichiometric action of Rad55–Rad57 eliminates the possibility that Rad55–Rad57 inhibition functions by binding Srs2 in solution. Rad55–Rad57 only slightly inhibit Srs2 helicase in the absence of Rad51 (Fig. 4g and Supplementary Fig. 10c). Control experiments show that this effect depends on Srs2 translocating in the expected 3' to 5' direction (Supplementary Fig. 10d), showing that Rad55–Rad57 inhibit Srs2 translocation on DNA to increase filament stability (Figs 1 and 2) and function (Supplementary Fig. 11). Direct visualization of human RAD51 filaments revealed that RAD51 is only able to form discontinuous short clusters on double-stranded DNA, as a result of frequent nucleation but limited extension^{18,19}. If this property holds

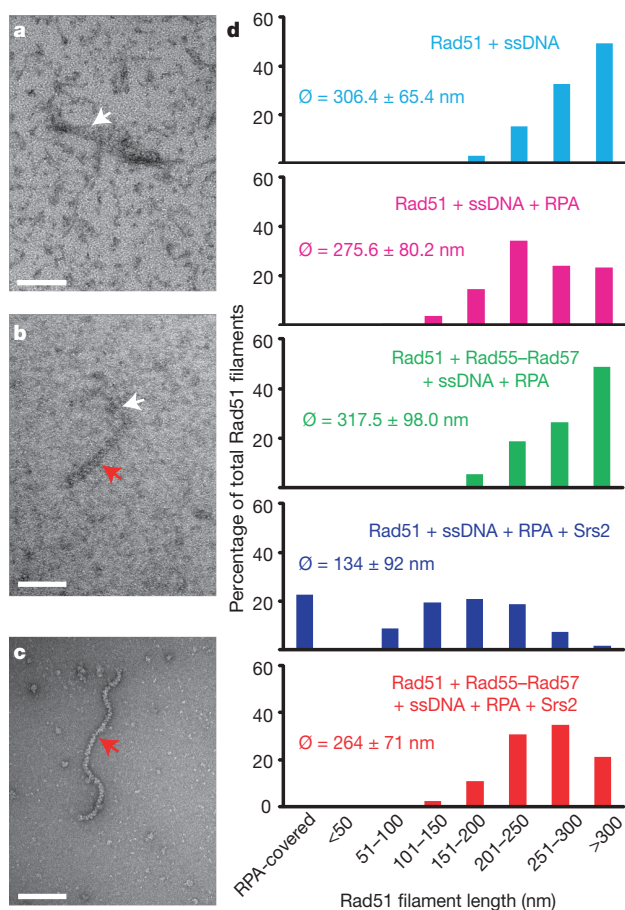


Figure 3 | Rad55–Rad57 inhibit disruption of Rad51 presynaptic filaments by Srs2. **a**, RPA–ssDNA complex. **b**, Short (145 nm) Rad51–ssDNA filament. **c**, Long (350 nm) Rad51–ssDNA filaments. **d**, Quantification of electron microscopic analysis. For each reaction condition 300–400 filaments were analysed (2.34 μ M Rad51, 7 μ M 600-nucleotide ssDNA, \pm 0.43 μ M Rad55–Rad57, \pm 0.21 μ M RPA, \pm 0.4 μ M Srs2), and the means ($\bar{\phi}$) \pm 1 s.d. and distributions of filament length classes are shown. Scale bars, 100 nm. White arrows indicate RPA–ssDNA complexes and red arrows indicate Rad51 filaments.

true for ssDNA, the formation of a co-filament with Rad51 by Rad55–Rad57 might provide a mechanism to form extended Rad51 filaments. This could also explain the increase in Rad55–Rad57 focus intensity over time after ionizing radiation exposure, and is consistent with the dependence of Rad55–Rad57 foci on Rad51 (ref. 10).

Our biochemical data are consistent with a model (Supplementary Fig. 1) whereby Rad51 presynaptic filament formation is modulated by a balance between the stabilizing function of Rad55–Rad57 and the destabilizing function of Srs2 antirecombinase. This model predicts that a deletion of *SRS2* should suppress the phenotypes caused by defects in Rad55–Rad57. In fact, *srs2Δ* completely suppresses the ionizing radiation sensitivity of *rad57* and *rad55* mutations in quantitative survival assays (Supplementary Fig. 12), consistent with semiquantitative results using *rad57* (ref. 20). However, *srs2Δ* only mildly suppresses the methyl methanesulphonate sensitivity (Supplementary Fig. 13) and recombination defect (Supplementary Fig. 14) of a *rad55* mutation, consistent with previous *rad57* data²⁰. The difference in suppression is likely related to a difference in substrates: ionizing radiation-induced DNA damage requires primarily double-strand break repair, whereas methyl

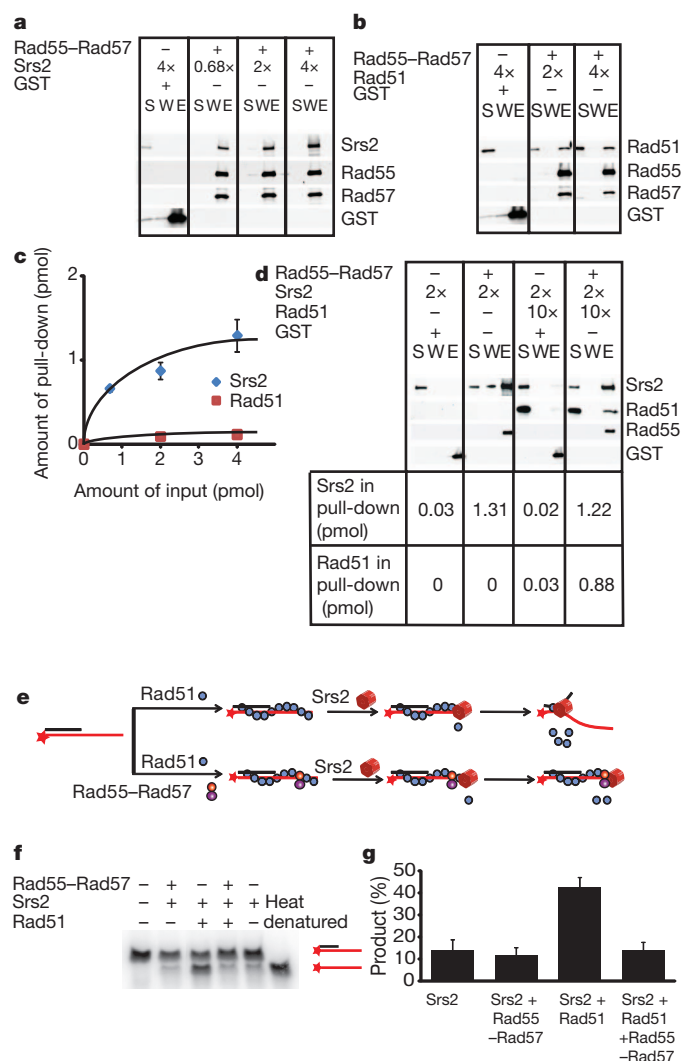


Figure 4 | Rad55–Rad57 interact with Srs2 and inhibit Srs2 helicase. **a**, Pull-down with 4 nM (1 pmol) Rad55–Rad57 and 2.7, 8, or 16 nM Srs2. **b**, Pull-down with 4 nM Rad55–Rad57 and 8 or 16 nM Rad51. **c**, Quantification of results in **a** and **b** and additional experiments. **d**, Pull-down with 4 nM Rad55–Rad57 and 8 nM Srs2 \pm 40 nM Rad51. GST was used as control. S, supernatant; W, wash; E, eluate. **e**, Helicase assay. **f**, Rad51 (28 nM) with or without Rad55–Rad57 (25 nM) were incubated with 1.5 nM 3'-tailed substrate before addition of 120 nM Srs2 protein. **g**, Product yields at 20 min. Means \pm 1 s.d., $n = 3$, are shown.

methanesulphonate-induced DNA damage and sister chromatid recombination require gap repair (Supplementary Fig. 1). We propose that the Rad51 presynaptic filament is a meta-stable reversible intermediate, whose dynamics in yeast are partially controlled by the balance of the filament-stabilizing activity of Rad55–Rad57 and the filament-destabilizing activity of the Srs2 helicase (Supplementary Fig. 1). This balance is likely to be influenced by the multiple post-translational modifications that have been identified to regulate Rad55–Rad57 (ref. 21) and Srs2 (ref. 22) functions (Supplementary Fig. 1). Together with the local availability of SUMO–PCNA, which specifically recruits Srs2 (refs 23–25), post-translational modifications may determine the balance between recombination and antirecombination in wild-type cells and explain the various degrees of suppression observed in the *srs2 rad55* (*rad57*) double mutants that depend on the type of DNA damage or genetic endpoint (double-strand break versus replication-fork-associated gap in Supplementary Fig. 1).

The human RAD51 paralogues have important roles in tumour suppression and human disease^{3,26}. Our studies established an unprecedented mechanism of anti-antirecombination that may serve as a

paradigm for the mechanism of action of the five human RAD51 paralogues. The diversification of the human RAD51 paralogues may reflect the multiplicity of human motor proteins that may disrupt RAD51 presynaptic filaments, including the RecQ-like helicases BLM and RECQL5 as well as FBH1 and FANCI^{27–30} or indicate additional functions during recombinational repair.

METHODS SUMMARY

Purification of yeast Rad51, Rad55–Rad57, RPA and Srs2, the biochemical assays and the electron microscopy analysis are detailed in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 September 2009; accepted 30 August 2011.

Published online 23 October 2011.

- Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
- Symington, L. S. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).
- Thacker, J. The RAD51 gene family, genetic instability and cancer. *Cancer Lett.* **219**, 125–135 (2005).
- Sung, P. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev.* **11**, 1111–1121 (1997).
- Lovett, S. T. & Mortimer, R. K. Characterization of null mutants of the RAD55 gene of *Saccharomyces cerevisiae*: effects of temperature, osmotic strength and mating type. *Genetics* **116**, 547–553 (1987).
- Beernink, H. T. H. & Morrical, S. W. RMPs: recombination/replication mediator proteins. *Trends Biochem. Sci.* **24**, 385–389 (1999).
- Morimatsu, K. & Kowalczykowski, S. C. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol. Cell* **11**, 1337–1347 (2003).
- Liu, J., Doty, T., Gibson, B. & Heyer, W. D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. *Nature Struct. Mol. Biol.* **17**, 1260–1262 (2010).
- Jensen, R. B., Carreira, A. & Kowalczykowski, S. C. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* **467**, 678–683 (2010).
- Lisby, M., Barlow, J. H., Burgess, R. C. & Rothstein, R. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* **118**, 699–713 (2004).
- Schiestl, R. H., Prakash, S. & Prakash, L. The SRS2 suppressor of *rad6* mutations of *Saccharomyces cerevisiae* acts by channeling DNA lesions into the RAD52 DNA repair pathway. *Genetics* **124**, 817–831 (1990).
- Aboussekhra, A. et al. RADH, a gene of *Saccharomyces cerevisiae* encoding a putative DNA helicase involved in DNA repair. Characteristics of *radH* mutants and sequence of the gene. *Nucleic Acids Res.* **17**, 7211–7219 (1989).
- Aguilera, A. & Klein, H. L. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics* **119**, 779–790 (1988).
- Krejci, L. et al. DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
- Veaute, X. et al. The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
- Antony, E. et al. Srs2 disassembles Rad51 filaments by a protein-protein interaction triggering ATP turnover and dissociation of Rad51 from DNA. *Mol. Cell* **35**, 105–115 (2009).
- Dupaigne, P. et al. The Srs2 helicase activity is stimulated by Rad51 filaments on dsDNA: implications for crossover incidence during mitotic recombination. *Mol. Cell* **29**, 243–254 (2008).
- Hilario, J., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules. *Proc. Natl Acad. Sci. USA* **106**, 361–368 (2009).
- Modesti, M. et al. Fluorescent human RAD51 reveals multiple nucleation sites and filament segments tightly associated along a single DNA molecule. *Structure* **15**, 599–609 (2007).
- Fung, C. W., Mozlin, A. M. & Symington, L. S. Suppression of the double-strand-break-repair defect of the *Saccharomyces cerevisiae* rad57 mutant. *Genetics* **181**, 1195–1206 (2009).
- Herzberg, K. et al. Phosphorylation of Rad55 on serines 2, 8, and 14 is required for efficient homologous recombination in the recovery of stalled replication forks. *Mol. Cell. Biol.* **26**, 8396–8409 (2006).
- Saponaro, M. et al. Cdk1 targets Srs2 to complete synthesis-dependent strand annealing and to promote recombinational repair. *PLoS Genet.* **6**, e1000858 (2010).
- Papouli, E. et al. Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol. Cell* **19**, 123–133 (2005).
- Pfander, B., Moldovan, G. L., Sacher, M., Hoege, C. & Jentsch, S. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **436**, 428–433 (2005).
- Burgess, R. C. et al. Localization of recombination proteins and Srs2 reveals anti-recombinase function *in vivo*. *J. Cell Biol.* **185**, 969–981 (2009).
- Meindl, A. et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nature Genet.* **42**, 410–414 (2010).
- Hu, Y. et al. RECQL5/Recq15 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev.* **21**, 3073–3084 (2007).
- Bugreev, D. V., Yu, X., Egelman, E. H. & Mazin, A. V. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. *Genes Dev.* **21**, 3085–3094 (2007).
- Sommers, J. A. et al. FANCI uses its motor ATPase to destabilize protein-DNA complexes, unwind triplexes, and inhibit RAD51 strand exchange. *J. Biol. Chem.* **284**, 7505–7517 (2009).
- Fugger, K. et al. Human Fbh1 helicase contributes to genome maintenance via pro- and anti-recombinase activities. *J. Cell Biol.* **186**, 655–663 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Alexeeva for the cell culture support. We thank P. Sung, R. Kolodner and L. Symington for plasmids and yeast strains. We are grateful to S. Kowalczykowski, N. Hunter, D. Castaño-Diez, P. Ringler and all members of the Heyer laboratory for discussions and comments on the manuscript. This work was supported by Postdoctoral Fellowship 17FT-0046 from the Tobacco-Related Disease Research Program (J.L.), by the European Community (LSHG-CT-2003-503303) and the Centre National de la Recherche Scientifique, the Commissariat à l'Energie Atomique (X.V., F.F.), by SystemsX.ch (H.S.), and the National Institutes of Health grants U54GM74929 (H.S.), CA92267 and GM58015 (W.-D.H.).

Author Contributions J.L. designed, performed and analysed all experiments, except the ionizing radiation survival assay, and helped write the manuscript. L.R. helped with the electron microscopy image collection and data analysis. X.V. purified the Srs2 protein. F.F. performed the ionizing radiation experiment. H.S. advised on the electron microscopy analysis. W.-D.H. conceived the project, designed experiments, coordinated collaborations, contributed to data analysis and wrote the manuscript. All authors discussed results and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.-D.H. (wdheyer@ucdavis.edu).

METHODS

Protein purification. Yeast Rad51, RPA and Srs2 proteins were purified as described^{15,31}. The purification of Rad55–Rad57 was adapted from a previously published protocol³². Yeast cells overexpressing GST–Rad55–His₆–Rad57 were grown and harvested as described³². Cells were disrupted in Buffer B containing 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 2.7 mM KCl, 1 M NaCl, 10% (v/v) glycerol, 10 mM β-mercaptoethanol and protease inhibitor cocktail (1 mM PMSF, 2 μM leupeptin, 1 μM pepstatin A and 1 mM benzamide) using glass beads (0.5 mm glass beads; BioSpec Products, Inc.). The cell lysate was centrifuged at 40,000 r.p.m. for 45 min using a Ti50.2 rotor. The supernatant was collected and loaded onto a pre-equilibrated Glutathione Sepharose 4B column (GE Healthcare). After washing with buffer A (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 M NaCl, 5 mM β-mercaptoethanol and 10% glycerol), the GST-tagged proteins were eluted with Buffer A containing 20 mM reduced glutathione plus protease inhibitor cocktail. Fractions containing the GST–Rad55–His₆–Rad57 heterodimer, as determined by 10% SDS–PAGE, were pooled and dialysed against Buffer C (50 mM NaH₂PO₄ pH 8.0, 1 M NaCl and 10% glycerol) containing the protease inhibitor cocktail. Then the pool was loaded onto a pre-equilibrated Ni-NTA agarose column and washed with Buffer C plus protease inhibitor cocktail. The bound complexes were eluted with Buffer C containing 0.5 M NaCl, 0.1 mM PMSF and 250 mM imidazole, and analysed by 10% SDS–PAGE. Fractions containing stoichiometric Rad55–Rad57 heterodimer were pooled, concentrated, dialysed into the storage buffer containing 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 0.1 mM EDTA, 1 mM DTT and 10% glycerol, and then stored in aliquots at –80 °C. The absence of contaminating enzymatic activities and DNA in protein preparations was verified as described³³.

Purification of 600-nucleotide ssDNA. As published previously¹⁷, 600-bp dsDNA fragments biotinylated at a one 5' end were generated by PCR from PstI-linearized φX174 DNA using primers WDHY427 5'-TTATCGAAGCGCGCATAAAT-3' and 5' biotinylated WDHY431 5'-GTCCTTCATTTCCATGCGG TG-3'. The biotinylated dsDNA was loaded onto a HiTrap Streptavidin HP column (Amersham Biosciences), and non-biotinylated single-stranded DNA was eluted with 60 mM NaOH.

Rad51–ssDNA filament assembly assay. In Fig. 1b, Rad51 (0.67 μM) was incubated with 4 μM ssDNA, in the presence or absence of 0.11 μM Rad55–Rad57, in buffer R containing 20 mM triethanolamine pH 7.5, 4 mM magnesium acetate, 2.5 mM ATP, 25 μg ml^{–1} BSA, 1 mM DTT, 90 mM NaCl and 5% glycerol for 10 min. Then 0.25% glutaraldehyde was used to crosslink the protein–DNA complexes for 15 min. The complexes were separated on a 0.5% agarose gel, stained with SYBR Gold, transferred to nitrocellulose membrane, and blotted with rabbit polyclonal anti-Rad51 or anti-Rad55 antibodies.

Rad51–ssDNA filament salt challenge assay. In Supplementary Fig. 3, Rad51 (0.267 μM) was incubated with 4 μM ssDNA, in the presence or absence of 0.067 μM Rad55–Rad57, in buffer R containing 20 mM triethanolamine pH 7.5, 4 mM magnesium acetate, 2.5 mM ATP, 25 μg ml^{–1} BSA, 1 mM DTT and 5% glycerol for 10 min. Then 5 M stock NaCl solution was added to the reaction to reach a final concentration of 500 mM for a further incubation of 30 min. Glutaraldehyde (0.25%) was used to crosslink the protein–DNA complexes for 15 min. Complexes were separated on a 0.5% agarose gel and stained with SYBR Gold. Proteins were transferred to nitrocellulose membrane and blotted with anti-Rad51 antibodies. All DNA concentrations refer to nucleotides (ssDNA) or base pairs (dsDNA).

Protein binding to ssDNA immobilized on magnetic beads. In Fig. 2b, a 5'-biotinylated oligonucleotide was immobilized onto magnetic streptavidin beads as previously described³⁴. The oligo sequence is 5'-CCCCCCCCCCCCCAAGATAATTTTTCGACTCATCAGAAATATCCGAAAGTGTTAACTTCTGCGTCATGGAAGCGATAAACTC-3'. In experiments containing Srs2, 10-μl slurry of beads containing 3 μM ssDNA was incubated with 1 μM Rad51 in the presence and absence of 0.1 μM Rad55–Rad57 in buffer containing 20 mM triethanolamine, 5 mM magnesium acetate, 4 mM ATP, 25 μg ml^{–1} BSA, 1 mM DTT, 5% glycerol and 25 mM NaCl for 10 min at 22 °C. Then 0.1 or 0.33 μM Srs2 protein was added and further incubated for 10 min. The beads were washed, and bound proteins were eluted and quantified as described³⁵. Background protein binding was typically less than 3%.

Topology-based assay for Rad51 dissociation. A published protocol was modified slightly for this assay using M13mp18 ssDNA³⁶. In Supplementary Fig. 4, 375 nM Rad51, with 0, 80 and 120 nM Rad55–Rad57, were incubated with 9 μM circular M13mp18 ssDNA in 25 μl of buffer containing 20 mM triethanolamine pH 7.5, 4 mM magnesium acetate, 25 μg ml^{–1} BSA, 1 mM DTT, and an ATP-regenerating system consisting of 2.5 mM ATP, 20 U ml^{–1} creatine kinase, and 20 mM creatine phosphate for 10 min at 30 °C. Then, 100 nM Srs2 and 150 nM RPA were added and incubated for 10 min, before the addition of topologically relaxed pUC19 dsDNA (7 μM in base pairs) and wheat germ DNA topoisomerase I (3 U). After

another 10 min incubation, reactions were stopped by addition of 4 μl stop solution consisting of 1% (w/v) SDS, 75 mM EDTA, 10 mg ml^{–1} protease K and further 30 min incubation at 37 °C. DNA species were resolved by electrophoresis on a 1% TBE-agarose gel and visualized using ultraviolet transillumination after ethidium bromide staining. The results were quantified using ImageQuant.

Protein interaction assays. GST–Rad55–His₆–Rad57 (4 nM) or 30 nM GST (GE Healthcare) were incubated with increasing amounts of either Srs2 or Rad51 in buffer P containing 25 mM Tris-HCl (pH 7.5), 10 mM magnesium acetate, 50 mM NaCl, 1 mM DTT, 10% glycerol and 0.05% NP-40 for 1 h at room temperature (Fig. 4a–d). Equilibrated and BSA-treated Glutathione-Sepharose 4B beads were added to the mixture and incubated for 1 h. The beads and supernatant were separated by centrifugation and the beads were washed twice with binding buffer P. The pulled-down protein complexes were eluted by boiling at 95 °C for 3 min in 10 μl SDS–PAGE loading buffer, separated through a 10% SDS–PAGE gel, and the protein bands were visualized through immunoblots and quantified by ImageQuant. In Fig. 4a, b, 1/16th of the supernatant and wash were loaded. In Fig. 4d, 1/7th of the supernatant and wash were loaded. For the competition protein binding assay (Supplementary Figs 7 and 8), the two proteins were incubated for 30 min before the addition of an increasing amount of the third challenging protein, as specified in the diagrams. After another 30-min incubation, equilibrated and BSA-treated Glutathione-Sepharose 4B beads were added to the mixture and incubated for 1 h. Analysis and quantification was performed as described above. The anti-Rad51, -Rad55 and -Rad57 antibodies were generated in rabbits, the anti-Srs2 antibody was purchased from Santa Cruz Biotechnology.

Helicase assay. The assay followed a published protocol and the substrates were prepared exactly as described before¹⁵. In Fig. 4f, 28 nM Rad51 with 0 or 25 nM Rad55–Rad57 were incubated with 1.5 nM oligo substrate with 3' tail in buffer containing 20 mM triethanolamine pH 7.5, 4 mM magnesium acetate, 25 μg ml^{–1} BSA, an ATP-regenerating system consisting of 2.5 mM ATP, 20 U ml^{–1} creatine kinase and 20 mM creatine phosphate, as well as either 1 mM DTT and 40 mM NaCl (Fig. 4e, g and Supplementary Fig. 10c, d) or 5 mM DTT and 10 mM NaCl (Supplementary Fig. 10a) for 10 min at 30 °C. Then 120 nM Srs2 protein was added to initiate the helicase reaction. After 20 min incubation, the reactions were stopped by adding 4.5 μl stop buffer containing 150 mM EDTA, 2% SDS, 163 nM unlabelled oligo, and 4.3 mg ml^{–1} protease K into 9 μl reaction sample. The DNA species were separated through electrophoresis on a 10% TBE–PAGE gel, which was dried and analysed by a Storm phosphorimager. The bands were quantified by densitometry using ImageQuant.

DNA strand exchange assay. In Supplementary Fig. 11, Rad51 (3.3 μM) was incubated with 0.3 μM Rad55–Rad57 or the corresponding amount of Rad55–Rad57 storage buffer and 10 μM φX174 ssDNA for 15 min at 30 °C in buffer containing 30 mM Tris-acetate (pH 7.5), 4 mM magnesium acetate, 75 mM NaCl, 1 mM DTT, 2.5 mM ATP, 50 μg ml^{–1} BSA, 20 mM phosphocreatine and 80 ng μl^{–1} creatine kinase. 0.56 μM RPA and 0, 333, 222, 167, 125 nM of Srs2 were added, and incubated for another 30 min. Then 10 μM (bp) PstI-linearized φX174 dsDNA and 4.8 mM spermidine were added and further incubated for 120 min. Samples were deproteinized and separated by electrophoresis on a 0.8% TBE-agarose gel. Images were recorded using a FluorChem8900 imaging system (Alpha Innotech) after staining with SYBR-Gold (Invitrogen), and quantified with ImageQuant. Percentage of joint molecule (JM) was calculated according to the equation $JM\% = (JM/1.5)/(JM/1.5 + NC + dsDNA)$. Percentage of product formation was calculated according to the equation $product\% = (JM/1.5 + NC)/(JM/1.5 + NC + dsDNA)$. NC, nicked circle.

Electron microscopy. To assemble the protein–DNA filament, 2.34 μM Rad51 protein, in the presence or absence of 0.43 μM Rad55–Rad57, was incubated with 7 μM 600-nt ssDNA (+) strand for 10 min at 30 °C in 20 mM triethanolamine pH 7.5, 4 mM magnesium acetate, 1 mM DTT and 3 mM ATP. RPA (0.21 μM) was added and incubated for another 10 min. Lastly, 0.4 μM Srs2 or buffer control was added and incubated for 10 min. The reaction mixtures were diluted 20-fold in 10 mM Tris-HCl pH 7.5, 50 mM NaCl and 5 mM MgCl₂ without chemical fixation. The samples were adsorbed onto 400 mesh carbon-coated copper grids (Ted Pella), negatively stained with 2% (w/v) uranyl acetate, blotted, and air-dried. Grids were imaged in a JEOL JEM-1230 transmission electron microscope (JEOL). Images of Rad51-filaments were randomly collected from different areas on the grid. 6–10 grids were used for each condition. Images were recorded at a nominal magnification of ×40,000 under minimum dose procedures on a Tietz 2,048 × 2,048 pixel CCD camera (TVIPS, Germany). Immunoaffinity gold labelling of GST–Rad55, as shown in Fig. 1d, was adapted from a published protocol³⁷. In brief, Rad51–Rad55–Rad57–ssDNA complexes were assembled as described above and crosslinked with 0.25% glutaraldehyde for 20 min, before deposition on grids. Grids were blocked in 50 μg ml^{–1} BSA in TBST for 30 min, and then incubated with goat anti-GST antibody (GE Healthcare) for 30 min. After three 5 min washes with 50 μg ml^{–1} BSA in TBST, the grids were incubated in TBST plus a 1:5

dilution of gold particles dressed with rabbit anti-goat antibody (BioAssay Works). After two 5 min washes in 50 $\mu\text{g ml}^{-1}$ BSA in TBST and one 5 min wash in 5 mM magnesium acetate, grids were stained with 2% (w/v) uranyl acetate before imaging.

***Saccharomyces cerevisiae* strains.** Strains used are listed in Supplementary Table 2.

Recombination assay. Spontaneous recombination rates between direct repeats were determined following a published fluctuation analysis protocol using the method of the median^{38,39}. The direct-repeat recombination substrate has two different *ade2* alleles separated by plasmid sequences and the *URA3* gene²⁰. Yeast strains were grown on YPD plates for 2 days at 30 °C for single colonies. For each strain, nine independent single colonies were randomly chosen and the entire colony was used to inoculate 4 ml YPD liquid culture. Liquid cultures were grown for 2–3 days at 30 °C to reach stationary phase. Cells were collected, washed with sterile H₂O, and suspended into 1 ml sterile H₂O. 100 μl of appropriate dilutions of each culture were spread on two plates each of SD-ADE-URA. Cells were incubated for 2 days at 30 °C. For each culture, the number of colonies on YPD were counted and totalled to determine the total cell number. The number of colonies on SD-ADE-URA were counted to determine the median number of recombinants. For each strain, recombination rates were measured independently three times and the mean values with standard deviations are shown.

MMS sensitivity assay. Yeast strains were grown overnight in liquid YPD to mid-log phase at 30 °C, and then diluted to OD_{600 nm} = 1. Serial dilutions of these cell cultures were made with sterile H₂O and spotted onto YPD plates with or without methyl methanesulphonate. Plates were incubated for 3 days at 30 °C or 5 days at 22 °C before photographing using a FluorChem8900 imaging system (Alpha Innotech).

Ionizing radiation survival assay. Exponentially growing cells (1×10^7 to 2×10^7 per ml) in YPD medium at 28 °C were collected by centrifugation, washed

in cold saline (0.9% NaCl), sonicated and resuspended in saline at the desired concentration. The cell suspension was γ -irradiated in a ¹³⁷Cs irradiator delivering 20 Gy min⁻¹. Aliquots of appropriate dilutions were spread on YPD-containing plates pre-warmed at either 23 °C or 34 °C. The plates were incubated at the corresponding temperature for 4 days (34 °C) or 6 days (23 °C) before counting the colonies. Platings were done in duplicate. The experiments were repeated at least three times, and the result of one typical assay is shown.

31. Solinger, J. A., Kiianitsa, K. & Heyer, W.-D. Rad54, a Swi2/Snf2-like recombinational repair protein, disassembles Rad51:dsDNA filaments. *Mol. Cell* **10**, 1175–1188 (2002).
32. Bashkurov, V. I., Herzberg, K., Haghazari, E., Vlasenko, A. S. & Heyer, W. D. DNA-damage induced phosphorylation of Rad55 protein as a sentinel for DNA damage checkpoint activation in *S. cerevisiae*. *Methods Enzymol.* **409**, 166–182 (2006).
33. Zhang, X. P. & Heyer, W. D. Quality control of purified proteins involved in homologous recombination. *Methods Mol. Biol.* **745**, 329–343 (2011).
34. Mazin, A. V., Alexeev, A. A. & Kowalczykowski, S. C. A novel function of Rad54 protein - Stabilization of the Rad51 nucleoprotein filament. *J. Biol. Chem.* **278**, 14029–14036 (2003).
35. Wolner, B., van Komen, S., Sung, P. & Peterson, C. L. Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast. *Mol. Cell* **12**, 221–232 (2003).
36. Schwendener, S. et al. Physical interaction of RECQ5 helicase with RAD51 facilitates its anti-recombinase activity. *J. Biol. Chem.* **285**, 15739–15745 (2010).
37. Van Dyck, E., Hajibagheri, N. M., Stasiak, A. & West, S. C. Visualisation of human RAD52 protein and its complexes with hRad51 and DNA. *J. Mol. Biol.* **284**, 1027–1038 (1998).
38. Lea, D. E. & Coulson, C. A. The distribution of the numbers of mutants in bacterial populations. *J. Genet.* **49**, 264–285 (1949).
39. Spell, R. M. & Jinks-Robertson, S. Determination of mitotic recombination rates by fluctuation analysis in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **262**, 3–12 (2004).

The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre

Johannes Fritsch^{1*}, Patrick Scheerer^{2*}, Stefan Frielingsdorf¹, Sebastian Kroschinsky², Bärbel Friedrich¹, Oliver Lenz¹ & Christian M. T. Spahn^{2,3}

Hydrogenases are abundant enzymes that catalyse the reversible interconversion of H₂ into protons and electrons at high rates¹. Those hydrogenases maintaining their activity in the presence of O₂ are considered to be central to H₂-based technologies, such as enzymatic fuel cells and for light-driven H₂ production². Despite comprehensive genetic, biochemical, electrochemical and spectroscopic investigations^{3–8}, the molecular background allowing a structural interpretation of how the catalytic centre is protected from irreversible inactivation by O₂ has remained unclear. Here we present the crystal structure of an O₂-tolerant [NiFe]-hydrogenase from the aerobic H₂ oxidizer *Ralstonia eutropha* H16 at 1.5 Å resolution. The heterodimeric enzyme consists of a large subunit harbouring the catalytic centre in the H₂-reduced state and a small subunit containing an electron relay consisting of three different iron-sulphur clusters. The cluster proximal to the active site displays an unprecedented [4Fe-3S] structure and is coordinated by six cysteines. According to the current model, this cofactor operates as an electronic switch depending on the nature of the gas molecule approaching the active site. It serves as an electron acceptor in the course of H₂ oxidation and as an electron-delivering device upon O₂ attack at the active site. This dual function is supported by the capability of the novel iron-sulphur cluster to adopt three redox states at physiological redox potentials^{7–9}. The second structural feature is a network of extended water cavities that may act as a channel facilitating the removal of water produced at the [NiFe] active site. These discoveries will have an impact on the design of biological and chemical H₂-converting catalysts that are capable of cycling H₂ in air.

More than two billion years ago, ancient microbes exploited the reducing power of H₂ for their metabolism; until today H₂ provides a valuable energy source which is used by H₂-oxidizing uptake hydrogenases. The reverse process, that is, the liberation of H₂, serves as safety valve to eliminate excessive reducing power under anaerobic conditions. This proton reduction reaction is catalysed by the group of H₂-evolving hydrogenases. All hydrogenases use abundant transition metals such as nickel and iron for catalysis, contrary to man-made H₂-converting catalysts that predominantly rely on rare precious metals.

Among three phylogenetically distinct types of hydrogenases, two enzyme classes prevail in nature. According to the metal content of their active sites they are classified as nickel-iron ([NiFe]) and di-iron ([FeFe]) hydrogenases¹⁰. [FeFe]-hydrogenases are highly productive in H₂ evolution, but are irreversibly inactivated during catalysis by even trace amounts of O₂ (ref. 11). [NiFe]-hydrogenases, however, function usually in the direction of H₂ oxidation and are less sensitive to O₂. In most cases, O₂ reacts with the active site giving rise to a mixture of inactive states, denoted as Ni-A and Ni-B, depending on the nature of the oxygen ligand bridging the Ni and Fe atoms in the active site¹². Both inactive forms, however, can be reactivated under reducing

conditions. Enzymes in the Ni-B state reactivate rapidly, whereas the recalcitrant Ni-A state requires long-term reactivation that may occur exclusively *in vitro*^{13,14}. Consequently, a prerequisite for a hydrogenase to function *in vivo* in the presence of O₂ is the strict avoidance of the Ni-A form and a continuous removal of the oxygen species related to the Ni-B state. These features are present in a small group of [NiFe]-hydrogenases that are designed to operate in mixtures of H₂ and O₂ (ref. 15). Knallgas bacterium *Ralstonia eutropha* H16 harbours at least three [NiFe]-hydrogenases capable of oxidizing H₂ at atmospheric pO₂. The best-characterized enzyme is the heterodimeric membrane-bound [NiFe]-hydrogenase (MBH), which is attached to the periplasmic side of the cytoplasmic membrane and feeds the electrons derived from H₂ oxidation via a membrane-integral *b*-type cytochrome directly into the respiratory chain (Fig. 1)¹⁶. Recent studies suggested that the iron-sulphur (Fe-S) cluster in the proximal position to the [NiFe] active site of MBH significantly differs in its electronic structure and function from conventional [4Fe-4S] cubanes, which are usually located at the corresponding position of O₂-sensitive [NiFe]-hydrogenases^{4,7}. Moreover, experimental evidence revealed that this particular cluster

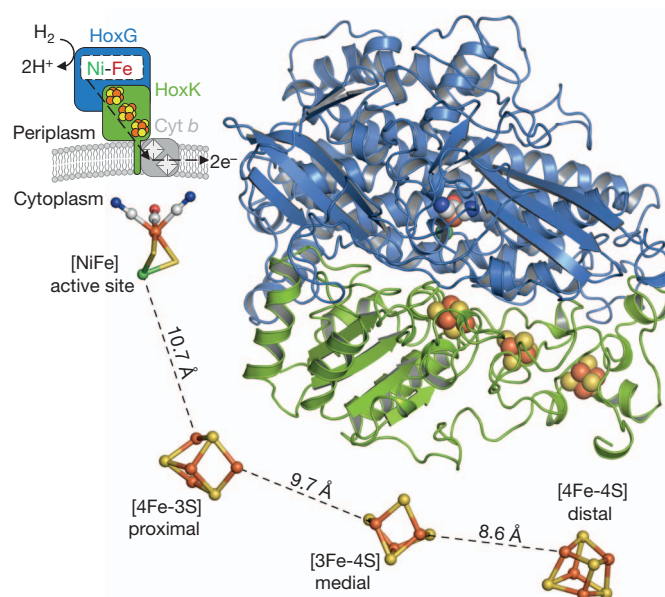


Figure 1 | Overall structure of the membrane-bound hydrogenase from *R. eutropha*. The upper-left inset is a cartoon depiction of the cellular localization of MBH. Electrons from H₂ oxidation are transferred through a relay of Fe-S clusters via a *b*-type cytochrome (Cyt *b*) to the respiratory chain. The ribbon representation shows the large (blue) and small (green) subunits of MBH. The catalytic centre and the three Fe-S clusters are symbolized as spheres. The spatial arrangement of MBH cofactors is illustrated in ball-and-stick representation in the lower part of the figure.

¹Mikrobiologie, Institut für Biologie, Humboldt-Universität zu Berlin, Chausseestraße 117, 10115 Berlin, Germany. ²Institut für Medizinische Physik und Biophysik (CC2), Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. ³Zentrum für Biophysik und Bioinformatik, Humboldt-Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany.

*These authors contributed equally to this work.

has a crucial role in the O₂ tolerance of MBH⁷. However, all structural interpretations of O₂ tolerance have been made on the basis of crystal structures from O₂-sensitive [NiFe]-hydrogenases^{12,17–20}.

To solve the crystal structure of MBH, the heterodimeric enzyme was purified from the solubilized membrane fraction of *R. eutropha* cells according to an optimized cultivation and purification protocol resulting in highly active and homogeneous protein preparations⁷. Dark-brown MBH crystals were harvested under a reducing atmosphere containing 5% H₂ and 95% N₂. The structure was solved by molecular replacement using the structure of the reduced hydrogenase of *Desulfovibrio vulgaris* Miyazaki F (Protein Data Bank accession 1WUL) as the search template, and the model was refined to a resolution of 1.5 Å.

The two MBH subunits show the typical overall topology of crystalized O₂-sensitive [NiFe]-hydrogenases (Fig. 1 and Supplementary Figs 1 and 2)^{12,17–20}. An initial omit electron density map readily uncovered the catalytic centre in the large subunit and the three Fe-S clusters in the small subunit (Supplementary Fig. 3). All four cofactors are spaced in distances of approximately 10 Å that allow electron transfer at physiologically relevant rates (Fig. 1). The bimetallic active site of MBH consists of a nickel atom coordinated by four cysteines, two of which are bridging ligands to the iron atom (Fig. 2a). Furthermore, the iron carries three diatomic ligands, one carbonyl (CO) and two cyanide (CN[−]) groups. The distance of 2.6 Å between the two metal atoms agrees with the reduced conformation of the active site¹². Consistent with previous spectroscopic observations^{4,7}, the first coordination sphere of the MBH catalytic centre is very similar to that of O₂-sensitive [NiFe]-hydrogenases (Supplementary Fig. 4). This indicates that the O₂ tolerance of MBH does not rely on a significant modification of the catalytic centre⁷.

Three Fe-S clusters in the MBH small subunit conduct the electron flow between the [NiFe]-active site and the cytochrome *b* (Fig. 1). The distal cluster relative to the active site is a cuboidal [4Fe-4S] centre coordinated by three cysteines and one histidine (Fig. 2d). It is shielded

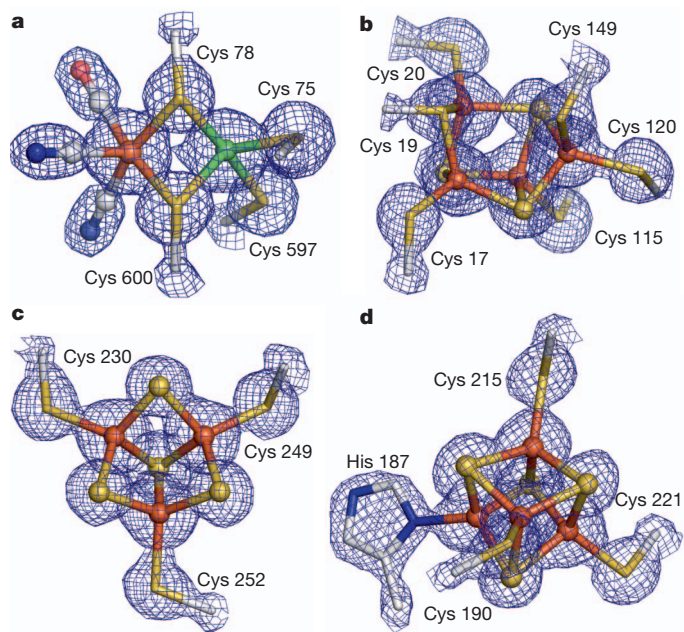


Figure 2 | Metal cofactors of the MBH. The blue meshes represent $2F_o - F_c$ electron densities contoured at 2.0σ . **a–d**, The $2F_o - F_c$ electron densities perfectly fit with the [(Cys)₂Ni(μ-Cys)₂Fe(CN)₂(CO)] centre (**a**), a proximal [4Fe-3S] cluster coordinated by six cysteine-derived sulphurs (**b**), a medial [3Fe-4S] cluster coordinated by three cysteine residues (**c**), and a distal [4Fe-4S] cluster coordinated by three cysteine residues and one histidine (**d**). All cofactors are shown in ball-and-stick representation; the coordinating amino acid side chains are depicted as stick models.

from the solvent by a 3₁₀-helix representing the first part of the carboxy-terminal α-helical extension of HoxK (Fig. 1 and Supplementary Fig. 1), which is essential for both anchoring the hydrogenase to the membrane and a tight connection to cytochrome *b*¹⁶. Three cysteine residues are involved in coordination of a [3Fe-4S] cluster which occupies the medial position, as observed in other [NiFe]-hydrogenases^{17,18,20} (Fig. 2c). The most surprising feature of MBH was observed at the position proximal to the active site. Instead of a common [4Fe-4S] cluster, a novel type of Fe-S cluster was found comprising four iron atoms but only three sulphides (Fig. 2b and Supplementary Fig. 3c). The conformation of this unprecedented [4Fe-3S] cluster is maintained by four cysteines (Cys 17^S, Cys 20^S, Cys 115^S, Cys 149^S, where S indicates the small subunit), which are conserved in all [NiFe]-hydrogenases, and two additional cysteines (Cys 19^S, Cys 120^S) exclusively found in O₂-tolerant enzymes (Supplementary Fig. 5)^{6–8}. In fact, Cys 19^S and Cys 120^S ligate three of the four iron atoms resulting in an open distorted conformation of the cluster (Fig. 3). In a concerted manner, Cys 120^S and Cys 149^S withdraw Fe3 from the cuboidal structure resulting in enlarged Fe–Fe distances of 3.5 and 4.0 Å to Fe1 and Fe4, respectively (Fig. 3). Notably, the typical Fe–Fe distance in [4Fe-4S] and [3Fe-4S] clusters is 2.7 Å (refs 17, 18, 20). The position of the missing sulphide in the [4Fe-3S] cluster, compared to a [4Fe-4S] centre, is occupied by the thiolate sulphur of Cys 19^S (Fig. 3).

The [4Fe-3S] cluster is clearly distinct from distorted, partially damaged clusters found in crystal structures of some standard hydrogenases^{19,20}. Moreover, it is unique among Fe-S centres with unusual structures (Supplementary Fig. 6). Surprisingly, the MBH proximal cluster shares similarity with one half of the reduced P-cluster of nitrogenase²¹. The P-cluster resembles a tandem of two cuboidal 4Fe-3S modules bridged by a single sulphide²². The resulting [8Fe-7S] centre is coordinated by six cysteine residues, two of which serve as bridging ligands. Interestingly, one of the two 4Fe-3S modules was found in a highly distorted conformation, which is analogous to the structure of the [4Fe-3S] cluster in MBH (Fig. 3 and Supplementary Fig. 6).

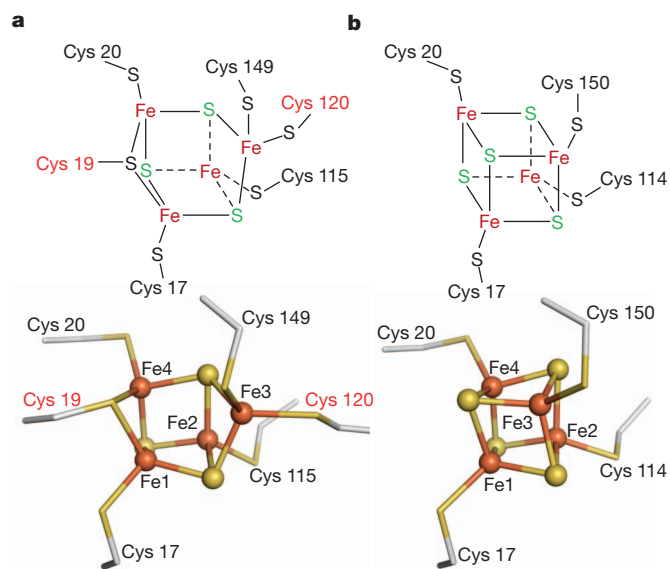


Figure 3 | Architecture of the proximal [4Fe-3S] cluster. **a**, Schematic model and structure of the [4Fe-3S] cluster and the corresponding coordinating cysteine ligands in the MBH small subunit. **b**, The proximal [4Fe-4S] cluster of the O₂-sensitive standard [NiFe]-hydrogenase from *D. vulgaris* Miyazaki F (Protein Data Bank accession 1WUL). In the upper schemes, the sulphides and cysteine-derived sulphur atoms are labelled in green and black, respectively. The cluster structures are shown in ball-and-stick representation; the coordinating amino acid side chains are depicted as stick models.

Oxygen tolerance implies that, upon approaching the catalytic centre, O_2 has to be removed reductively through an immediate delivery of four electrons and protons for the complete reduction of O_2 to water^{5,7}. Because the oxidized active site is blocked and cannot bind H_2 , electrons must be delivered by reverse electron flow^{3,5,7,16}. This feature seems to be linked to the previously determined, comparatively high redox potentials of the Fe-S clusters in MBH, which range from -180 mV to $+160$ mV. Moreover, the proximal cluster alone appears to undergo two redox transitions within an extraordinary narrow potential window from -60 mV to $+160$ mV (refs 4, 7, 9, 16, 23). The unique capability to carry two electrons at the same time at physiologically relevant potentials is in perfect agreement with the precisely assigned redox transitions mediated by the Fe-S clusters of the O_2 -tolerant, MBH-related hydrogenase I from *Aquifex aeolicus*⁸. Generally, the potential range between the $3+/2+$ and $2+/1+$ transitions of high-potential as well as ferredoxin-type $[4Fe-4S]$ clusters is approximately 1,000 mV, placing either of the transitions beyond physiological relevance^{24,25}. However, the corresponding potential window of the two redox transitions mediated by the $[4Fe-3S]$ cluster is only 220 mV (refs 8, 9, 23).

Compared to the rather symmetric $[4Fe-4S]$ clusters, the four iron atoms of the $[4Fe-3S]$ cluster are coordinated by a higher number of cysteine-derived thiolates and a lower number of sulphides (Fig. 3). This structural information in combination with the interpretation of data from interdisciplinary studies^{3-5,7} suggest that three out of four iron atoms reside formally in the $3+$ state in the most oxidized form of the $[4Fe-3S]$ cluster, which has been substantiated by Mössbauer experiments performed on hydrogenase I from *A. aeolicus*⁸. According to the current understanding^{26,27}, the increased thiolate:sulphide ratio elevates the redox potentials of all redox transitions mediated by an Fe-S cluster. In the case of the $[4Fe-3S]$ cluster, this would explain the high value of -60 mV of the low-potential redox transition, but certainly not the relatively low value of the high-potential redox transition ($E_M = +160$ mV). Conventional high-potential $[4Fe-4S]$ clusters are embedded in a hydrophobic pocket. Consequently, the number of hydrogen bonds (particularly those from water molecules) to sulphur ligands is low, which in turn leads to a high covalency of the Fe-S bonds and poises the $3+/2+$ transition to a physiological potential range²⁶⁻²⁹. A conserved water molecule in O_2 -sensitive hydrogenases is replaced by the Cys 120^S thiolate sulphur. However, two well-defined water molecules (Wat366, Wat447, Supplementary Fig. 3b) were found within hydrogen-bonding distance to the $[4Fe-3S]$ cluster. This leaves the question open as to how the protein environment tunes the high-potential transition of the proximal $[4Fe-3S]$ cluster. Notably, the open conformation of the $[4Fe-3S]$ cluster provides enhanced structural flexibility permitting redox-dependent rearrangements which have been observed, for example, for the P-cluster of nitrogenase²².

As discussed above, H_2 conversion in the presence of O_2 implies continuous production of H_2O at the active site^{5,7,16}. Thus, H_2O needs to be continuously removed from the protein core to the surface. It is rather unlikely that H_2O molecules escape through the proposed hydrophobic gas channels³⁰, which are also observed in MBH (Supplementary Fig. 7). The structure of MBH uncovered water-filled cavities that connect the active site with the solvent (Fig. 4) and are absent in the crystal structures of O_2 -sensitive $[NiFe]$ -hydrogenases (Supplementary Fig. 8). The additional cavity close to Cys 81^L (where L indicates the large subunit) in MBH originates from the replacement of a bulky tyrosine, present in most O_2 -sensitive hydrogenases (Supplementary Fig. 5), by the small Gly 80^L residue. Two gates on opposite sides of this water pocket seem to prevent unrestricted water/proton flow between the active site and the protein surface. The imidazol group of His 220^L, which is conserved in O_2 -tolerant hydrogenases and the presence of which strictly correlates with the occurrence of Gly 80^L, disrupts the direct connection to a water-filled extension of the gas channel that reaches the $[NiFe]$ centre (Fig. 4). The proposed water transfer to the surface is gated by salt bridges formed between

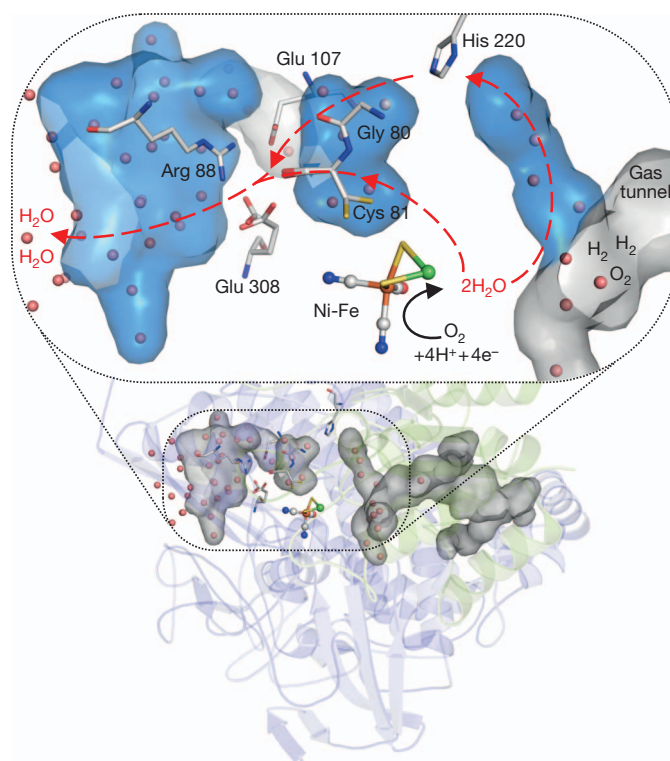


Figure 4 | Proposed water/proton transfer pathway from the MBH active site to the protein surface. Relevant water-filled cavities are coloured in blue; the water molecules are represented as red spheres. Other cavities including the mostly hydrophobic gas channel³⁰ are shown in grey. Amino acids in the large subunit involved in the proposed water/proton transfer (red arrows) are depicted as sticks. Cys 81^L and Glu 308^L are shown in both side-chain conformations (see also Supplementary Fig. 9). For the complete reduction of O_2 to two H_2O molecules, four electrons must be delivered rapidly by the electron relay of the small subunit (Supplementary Fig. 7).

Arg 88^L and the glutamate residues, Glu 107^L and Glu 308^L (Fig. 4). Remarkably, we identified two rotamers of Glu 308^L, indicating a transient opening of the gate (Supplementary Fig. 9). Thus, the structural data are compatible with a controlled translocation of water molecules, probably with accompanied proton transfer, supporting the model that O_2 -tolerant hydrogenases form water as a catalytic by-product during H_2 conversion in the presence of O_2 .

METHODS SUMMARY

Crystallization and structure determination of MBH. The membrane fraction containing *Strep*-tagged MBH was prepared from *R. eutropha* cells grown heterotrophically under microaerobic, hydrogenase-derepressing conditions⁷. After treatment with $K_3[Fe(CN)_6]$, membrane proteins were solubilized with Triton X-114, and MBH was purified via *Strep*-Tactin affinity chromatography. MBH crystals were obtained using sitting-drop vapour diffusion. On a microbridge, 8 μ l protein solution containing 0.1 mM MBH were mixed with 8 μ l of reservoir solution containing 20–30% PEG3350, and 0.1 M Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane, pH 5.5–6.5. Crystals of MBH grew to full size in the drops within 5 days. The crystals were directly frozen in liquid nitrogen for screening and X-ray analysis at the synchrotron ESRF, Grenoble (France).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April; accepted 23 August 2011.

Published online 16 October 2011.

1. Cammack, R., Frey, M. & Robson, R. *Hydrogen as a Fuel, Learning from Nature* (Taylor & Francis, 2001).
2. Friedrich, B., Fritsch, J. & Lenz, O. Oxygen-tolerant hydrogenases in hydrogen-based technologies. *Curr. Opin. Biotechnol.* **22**, 358–364 (2011).

3. Ludwig, M., Cracknell, J. A., Vincent, K. A., Armstrong, F. A. & Lenz, O. Oxygen-tolerant H_2 oxidation by membrane-bound [NiFe] hydrogenases of *Ralstonia* species. Coping with low level H_2 in air. *J. Biol. Chem.* **284**, 465–477 (2009).
4. Saggi, M. *et al.* Spectroscopic insights into the oxygen-tolerant membrane-associated [NiFe] hydrogenase of *Ralstonia eutropha* H16. *J. Biol. Chem.* **284**, 16264–16276 (2009).
5. Cracknell, J. A., Wait, A. F., Lenz, O., Friedrich, B. & Armstrong, F. A. A kinetic and thermodynamic understanding of O_2 tolerance in [NiFe]-hydrogenases. *Proc. Natl Acad. Sci. USA* **106**, 20681–20686 (2009).
6. Lukey, M. J. *et al.* How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J. Biol. Chem.* **285**, 3928–3938 (2010).
7. Goris, T. *et al.* A unique iron-sulfur cluster is crucial for oxygen tolerance of a [NiFe]-hydrogenase. *Nature Chem. Biol.* **7**, 310–318 (2011).
8. Pandelia, M. E. *et al.* Characterization of a unique [FeS] cluster in the electron transfer chain of the oxygen tolerant [NiFe] hydrogenase from *Aquifex aeolicus*. *Proc. Natl Acad. Sci. USA* **108**, 6097–6102 (2011).
9. Schneider, K., Patil, D. S. & Cammack, R. Electron-spin-resonance properties of membrane-bound hydrogenases from aerobic hydrogen bacteria. *Biochim. Biophys. Acta* **748**, 353–361 (1983).
10. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
11. Stripp, S. T. *et al.* How oxygen attacks [FeFe] hydrogenases from photosynthetic organisms. *Proc. Natl Acad. Sci. USA* **106**, 17331–17336 (2009).
12. Ogata, H., Lubitz, W. & Higuchi, Y. [NiFe] hydrogenases: structural and spectroscopic studies of the reaction mechanism. *Dalton Trans.* 7577–7587 (2009).
13. Vincent, K. A., Parkin, A. & Armstrong, F. A. Investigating and exploiting the electrocatalytic properties of hydrogenases. *Chem. Rev.* **107**, 4366–4413 (2007).
14. De Lacey, A. L., Fernandez, V. M., Rousset, M. & Cammack, R. Activation and inactivation of hydrogenase function and the catalytic cycle: spectroelectrochemical studies. *Chem. Rev.* **107**, 4304–4330 (2007).
15. Schwartz, E. & Friedrich, B. in *The Prokaryotes* (eds Dworkin, M. *et al.*) 496–563 (Springer, 2006).
16. Lenz, O. *et al.* H_2 conversion in the presence of O_2 as performed by the membrane-bound [NiFe]-hydrogenase of *Ralstonia eutropha*. *ChemPhysChem* **11**, 1107–1119 (2010).
17. Volbeda, A. *et al.* Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).
18. Volbeda, A. *et al.* Structural differences between the ready and unready oxidized states of [NiFe] hydrogenases. *J. Biol. Inorg. Chem.* **10**, 239–249 (2005).
19. Matias, P. M. *et al.* [NiFe] hydrogenase from *Desulfovibrio desulfuricans* ATCC 27774: gene sequencing, three-dimensional structure determination and refinement at 1.8 Å and modelling studies of its interaction with the tetraheme cytochrome c_3 . *J. Biol. Inorg. Chem.* **6**, 63–81 (2001).
20. Ogata, H., Kellers, P. & Lubitz, W. The crystal structure of the [NiFe] hydrogenase from the photosynthetic bacterium *Allochrochromatium vinosum*: characterization of the oxidized enzyme (Ni-A state). *J. Mol. Biol.* **402**, 428–444 (2010).
21. Seefeldt, L. C., Hoffman, B. M. & Dean, D. R. Mechanism of Mo-dependent nitrogenase. *Annu. Rev. Biochem.* **78**, 701–722 (2009).
22. Peters, J. W. *et al.* Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry* **36**, 1181–1187 (1997).
23. Knüttel, K. *et al.* Redox properties of the metal centers in the membrane-bound hydrogenase from *Alcaligenes eutrophus* CH34. *Bull. Polish Acad. Sci.* **42**, 495–511 (1994).
24. Cammack, R. “Super-reduction” of chromate high-potential iron-sulphur protein in the presence of dimethyl sulphoxide. *Biochem. Biophys. Res. Commun.* **54**, 548–554 (1973).
25. Thomson, A. J. *et al.* Low-temperature magnetic circular-dichroism evidence for the conversion of 4-iron-sulfur clusters in a ferredoxin from *Clostridium pasteurianum* into 3-iron-sulfur clusters. *Biochim. Biophys. Acta* **637**, 423–432 (1981).
26. Capozzi, F., Ciurli, S. & Luchinat, C. Coordination sphere versus protein environment as determinants of electronic and functional properties of iron-sulfur proteins. *Metal Sites Proteins Models* **90**, 127–160 (1998).
27. Carter, C. W. Jr. New stereochemical analogies between iron-sulfur electron transport proteins. *J. Biol. Chem.* **252**, 7802–7811 (1977).
28. Dey, A. *et al.* Solvent tuning of electrochemical potentials in the active sites of HiPIP versus ferredoxin. *Science* **318**, 1464–1468 (2007).
29. Heering, H. A., Bultink, B. M., Hagen, W. R. & Meyer, T. E. Influence of charge and polarity on the redox potentials of high-potential iron-sulfur proteins: evidence for the existence of two groups. *Biochemistry* **34**, 14675–14686 (1995).
30. Montet, Y. *et al.* Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nature Struct. Biol.* **4**, 523–526 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to U. Müller, M. Weiss and the scientific staff of the BESSY-MX/Helmholtz Zentrum Berlin für Materialien und Energie at beamlines BL 14.1 and BL 14.2, D. von Stetten and A. Royant of the ID29S-Cryobench (ESRF, Grenoble) and the European Synchrotron Radiation Facility (ESRF, Grenoble) at beamlines ID23-1, ID23-2, ID14-1 and ID 14-4, where the data were collected, for support. This work was supported by the EU/FP7 programme Solar-H2 (to J.F.), the DFG Cluster of Excellence ‘Unifying Concepts in Catalysis’ (to S.F., B.F., O.L.), and the Sfb740 (to C.M.T.S.). P.S. acknowledges K. P. Hofmann and his advanced investigator ERC grant (ERC-2009/249910—TUDOR) for support.

Author Contributions J.F. and P.S. are joint first authors. J.F. optimized cell growth conditions as well as the MBH purification procedure. P.S. conducted the crystallization screening; P.S., J.F. and S.F. optimized MBH crystallization conditions. P.S. and S.K. collected the X-ray diffraction data. P.S. performed data processing, solved and refined the MBH structure. B.F., O.L. and P.S. coordinated the project. J.F., P.S., S.F. and O.L. analysed data. J.F., P.S., S.F., B.F., O.L. and C.M.T.S. wrote the manuscript.

Author Information Atomic coordinates and structure factors for the reported structure have been deposited in the Protein Data Bank with the accession code 3RGW. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S. (patrick.scheerer@charite.de), O.L. (oliver.lenz@cms.hu-berlin.de) or C.M.T.S. (christian.spahn@charite.de).

METHODS

Media and cell growth conditions. Basic media and growth conditions for *R. eutropha* have been described previously^{7,31}. *Ralstonia eutropha* HF649 was cultivated heterotrophically at 30 °C in a modified FGN mineral medium containing 0.04% wt/vol fructose, 0.4% wt/vol glycerol and 40 µM FeCl₃. 4,000 ml cultures were shaken in baffled 5,000 ml Erlenmeyer flasks at 120 r.p.m. under air until they reached an optical density at 436 nm of 12 ± 1. Cells were harvested by centrifugation at 6,000g for 15 min at 4 °C.

Isolation of membranes and Strep-Tactin affinity chromatography. Fractionation and purification steps were performed at 4 °C. Cells were re-suspended in buffer (3 ml 65 mM potassium phosphate [K-PO₄], 300 mM NaCl, pH 7.0, per 1 g wet weight) containing Complete EDTA-free protease inhibitor cocktail (Roche Applied Science) and DNase I. The cell suspension was subsequently disrupted in a French pressure cell (SLM Aminco) via two passages at 1,241 bar. The resulting crude extract was treated by sonication (Branson Sonifier) and the cell debris was removed by low-speed centrifugation (4,000g, 30 min). K₃[Fe(CN)₆] was added to the crude extract at a final concentration of 50 mM. Membrane and soluble fractions were separated by ultracentrifugation (100,000g for 60 min). The membrane pellet was washed with re-suspension buffer containing Complete protease inhibitor cocktail (Roche Applied Science), followed by ultracentrifugation (100,000g for 50 min). Membrane proteins were solubilized in 10 ml buffer (65 mM K-PO₄, 300 mM NaCl, 2% wt/vol Triton X-114, Complete protease inhibitor cocktail, pH 7.02) per 1 g of membrane pellet by stirring on ice for 2 h. After ultracentrifugation (100,000g, 45 min), the supernatant containing the solubilized membrane extract was loaded onto Strep-Tactin Superflow columns (IBA) which were run by gravity flow. To remove unbound proteins, the columns were washed with 12 bed volumes of re-suspension buffer and proteins were eluted with buffer containing 50 mM K-PO₄, 150 mM NaCl, 5 mM desthiobiotin and 10% wt/vol glycerol at pH 7. MBH-containing fractions were pooled, concentrated and the buffer was exchanged to 40 mM K-PO₄, 150 mM NaCl, 10% wt/vol glycerol, pH 5.5, with a centrifugal filter device (Amicon Ultra-15 PL-30, Millipore). Protein concentrations were determined with the BCA-kit (Pierce) with bovine serum albumin as standard.

Hydrogenase activity assay. Spectrophotometric activity measurements of purified MBH were conducted at 30 °C in a rubber-stoppered cuvette containing H₂-saturated K-PO₄ buffer (50 mM, pH 5.5) and methylene blue as the electron acceptor³².

Crystallization. MBH was used for crystallization at concentrations up to 14 mg ml⁻¹ (7–14 mg ml⁻¹). Crystallization screens by the sparse matrix method³³ were carried out by the sitting-drop vapour diffusion method testing more than 1,000 crystallization conditions at 277 K and 291 K using 96-well MRC plates. Promising conditions were systematically screened further by changing protein concentration, pH and the concentration of precipitation agents. Optimized MBH crystals could be grown by sitting-drop vapour diffusion method at 282 K using 24-well Linbro plates. Each sitting drop was prepared on a microbridge by mixing equal volumes (8 µl each) of MBH and reservoir solution. The reservoir solution contained 20–30% polyethylene glycol 3350, 100 mM Bis-(2-hydroxy-ethyl)-amino-tris(hydroxymethyl)-methane buffer, pH 5.5–6.5. Dark-brown MBH crystals appeared within 1–2 days and grew further for 4–5 days and were harvested under an atmosphere composed of 5% H₂ and 95% N₂. MBH crystals were flash frozen in liquid nitrogen with (90% (v/v) reservoir solution and 10% (w/v) polyethylene glycol 400) and without further cryoprotection. Fully grown crystals had dimensions of approximately 1.4 × 0.3 × 0.3 mm³. Dissolved MBH crystals exhibited the same H₂ oxidation activity of 130 µmol H₂ per min per mg of protein as the protein before crystallization, demonstrating that the crystallization process preserved the catalytic activity of the enzyme.

Structure analysis. Diffraction data collection was performed at 100 K using synchrotron X-ray sources at ESRF, Grenoble, France, and BESSY II, Berlin, Germany. Best diffraction data were collected at beamline ID14-4 at ESRF, at λ = 0.9395 Å. The crystal to ADSC Q315r detector distance was fixed at 127.8 mm for MBH. The rotation increment for each frame was 0.5° with an exposure time of 1 s. To reduce significantly the radiation damage on the crystal we used the helical data collection protocol at beamline ID14-4³⁴. All images were

indexed, integrated and scaled using the XDS program package³⁵ and CCP4 program SCALA^{36,37}. Crystals belong to orthorhombic space group P2₁2₁2₁ (a = 73.09 Å, b = 95.65 Å, c = 119.15 Å, α = β = γ = 90°). Supplementary Table 1 summarizes the statistics for crystallographic data collection and structural refinement.

Initial phases for MBH in the H₂-reduced state were obtained by conventional molecular replacement protocol (rotation, translation, rigid body fitting) using the [NiFe]-hydrogenase structure of *Desulfovibrio vulgaris* (Protein Data Bank accession 1WUL) as the initial search model. Molecular replacement was achieved using the CCP4 program PHASER^{37,38} by first placing the MBH heterodimer (rotation function [RFZ]: Z = 21.7; translation function [TFZ]: Z = 36.2 for MBH; RFZ and TFZ as defined by PHASER). In subsequent steps, torsion angle molecular dynamics, simulated annealing using a slow-cooling protocol and a maximum likelihood target function, energy minimization, and B-factor refinement by the program CNS³⁹ were carried out in the resolution range 74.6–1.5 Å. After the first round of refinement, all [Fe_n-S_n] clusters and [NiFe]-active site were clearly visible in the electron density of both σA-weighted 2F_o – F_c maps, as well as in the σA-weighted simulated annealing omitted density maps (Supplementary Fig. 3). Restrained, individual B-factors were refined and the crystal structure was finalized by the CCP4 program REFMAC5 and CCP4^{37,40} and PHENIX⁴¹. The final model has agreement factors R_{free} and R_{cryst} of 15.2% and 13.9%. Manual rebuilding of the MBH model and electron density interpretation was performed after each refinement cycle using the program COOT⁴². Structure validation was performed with the programs PROCHECK⁴³ and WHAT_CHECK⁴⁴. Potential hydrogen bonds and van der Waals contacts were analysed using the programs HBPLUS⁴⁵ and LIGPLOT⁴⁶. All crystal structure superpositions of backbone α carbon traces were performed using the CCP4 program LSQKAB³⁷. The solvent-accessible area was calculated using the PISA Server⁴⁷. All molecular graphics representations were created using PyMol⁴⁸.

- Schubert, T., Lenz, O., Krause, E., Volkmer, R. & Friedrich, B. Chaperones specific for the membrane-bound [NiFe]-hydrogenase interact with the Tat signal peptide of the small subunit precursor in *Ralstonia eutropha* H16. *Mol. Microbiol.* **66**, 453–467 (2007).
- Schink, B. & Schlegel, H. G. The membrane-bound hydrogenase of *Alcaligenes eutrophus*. I. Solubilization, purification, and biochemical properties. *Biochim. Biophys. Acta* **567**, 315–324 (1979).
- Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409–411 (1991).
- Flot, D. et al. The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).
- Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *Appl. Cryst.* **26**, 283–291 (1993).
- Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature* **381**, 272 (1996).
- McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
- Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- DeLano, W. L. The PyMOL Molecular Graphics System (<http://www.pymol.org>) (2002).

Structural basis for a [4Fe–3S] cluster in the oxygen-tolerant membrane-bound [NiFe]–hydrogenase

Yasuhito Shomura^{1,2}, Ki-Seok Yoon^{3†}, Hirofumi Nishihara³ & Yoshiki Higuchi^{1,2,4}

Membrane-bound respiratory [NiFe]–hydrogenase (MBH), a H₂-uptake enzyme found in the periplasmic space of bacteria, catalyses the oxidation of dihydrogen: H₂ → 2H⁺ + 2e[−] (ref. 1). In contrast to the well-studied O₂-sensitive [NiFe]–hydrogenases (referred to as the standard enzymes), MBH has an O₂-tolerant H₂ oxidation activity^{2–4}; however, the mechanism of O₂ tolerance is unclear⁵. Here we report the crystal structures of *Hydrogenovibrio marinus* MBH in three different redox conditions at resolutions between 1.18 and 1.32 Å. We find that the proximal iron-sulphur (Fe-S) cluster of MBH has a [4Fe-3S] structure coordinated by six cysteine residues—in contrast to the [4Fe-4S] cubane structure coordinated by four cysteine residues found in the proximal Fe-S cluster of the standard enzymes—and that an amide nitrogen of the polypeptide backbone is deprotonated and additionally coordinates the cluster when chemically oxidized, thus stabilizing the superoxidized state of the cluster. The structure of MBH is very similar to that of the O₂-sensitive standard enzymes except for the proximal Fe-S cluster. Our results give a reasonable explanation why the O₂ tolerance of MBH is attributable to the unique proximal Fe-S cluster; we propose that the cluster is not only a component of the electron transfer for the catalytic cycle, but that it also donates two electrons and one proton crucial for the appropriate reduction of O₂ in preventing the formation of an unready, inactive state of the enzyme.

An increasing interest in exploring H₂ as a clean energy carrier has encouraged the use of hydrogenases as efficient catalysts for H₂ oxidation in fuel cells and for H₂ production. However, a substantial obstacle is the inhibition or inactivation of hydrogenases by O₂. For this reason, elucidating the molecular basis of the O₂ tolerance of MBH is crucial, and has implications for biotechnological development in this area.

The standard [NiFe]–hydrogenase and MBH are phylogenetically related heterodimeric enzymes, sharing ~45% amino acid sequence identity (Supplementary Table 1 and Supplementary Fig. 1a, b). Whereas the former is mostly found in microorganisms living in strictly anaerobic environments, the latter is found as a component of the aerobic respiratory system in some bacteria^{6–8}. The heterodimeric unit of MBH associates with a membrane-anchored *b*-type cytochrome subunit (Cyt *b*) and the inner membrane through a carboxy-terminal transmembrane helix in the small subunit, which is absent in the standard enzyme. The MBH–Cyt *b* complex catalyses the oxidation of H₂ at the periplasmic space to release protons and the reduction of quinones in the inner membrane to take up protons from the cytoplasmic side, so as to generate a proton gradient across the membrane.

The overall structure of the heterodimeric unit of MBH is similar to those of the standard enzymes^{9–12} (Supplementary Table 1), but we observed a dimer of heterodimers in an asymmetric unit of MBH crystals in which the transmembrane helices had been spontaneously cleaved off during purification (Fig. 1). In contrast to the heterodimeric structure of all of the standard enzymes reported so far¹³, we consider the heterotetrameric structure of MBH to be biologically relevant for

the following reasons: the contact area between the protomers occupies as much as 11% of the total surface area; and the C termini of the small subunits of the crystalline enzyme (which had been located near the Cyt *b* subunits *in vivo*) and the distal Fe-S clusters that transfer electrons to Cyt *b* face the same side of the complex.

Our structural analyses have shown that the proximal Fe-S cluster of MBH is not a [4Fe-4S] type as in the standard enzyme, but a [4Fe-3S] type, where one of the corner sulphides is missing and two additional cysteine thiolates coordinate iron atoms (Fig. 2a–d). The proposal made for MBH of *Ralstonia eutropha* H16 (based on the sequence alignment and spectroscopic analyses¹⁴) that the proximal Fe-S cluster has a structure and electronic properties distinct from those of the standard enzyme did not conflict with the X-ray structure elucidated in this study. Specifically, Sγ of Cys 25^S (where superscript S (or L) denotes that the residue is in the small (or large) subunit) replaces a sulphide that would have bound Fe1, Fe2 and Fe4 in the [4Fe-4S] cubane cluster to coordinate Fe1 and Fe2, and the Sγ of Cys 158^S coordinates Fe4 (Fig. 2a, b). Consequently, a bond between a sulphur and an iron atom in the standard [4Fe-4S] cluster is missing, but all four iron atoms are coordinated by four ligands. Despite the difference between the cluster species, the main-chain trace (that is, the Cα positions) around the proximal cluster of MBH is very similar to that of the standard enzyme (Fig. 3).

A structural comparison between different redox conditions revealed redox-dependent structural changes in the proximal Fe-S cluster. In the ferricyanide-oxidized condition, the amide nitrogen of Cys 26^S replaced the S3 ligand of Fe2, and the cluster became more asymmetric (Fig. 2c, d). Hereafter, the structures of the [4Fe-3S] clusters in the H₂-reduced and the ferricyanide-oxidized conditions will be termed the ‘symmetric’ and ‘asymmetric’ forms, respectively. Notably, the structures of the two forms were virtually identical except for the position of Fe2. The amide nitrogen in the asymmetric form is considered to be deprotonated, because the distance between the amide nitrogen and Fe2 is 2.09 Å, and C=O and N atoms in the amide and Fe2 are nearly coplanar. The deprotonation of the amide indicates that there is an additional negative charge at this amide/carbonyl, which should stabilize a higher oxidation state of the iron atoms in the asymmetric form, as described below.

The electron density map of the cluster in the air-oxidized condition revealed that the crystals contain both the symmetric and asymmetric forms (Supplementary Fig. 2). A considerable body of X-ray diffraction data and the structural refinement of different crystals obtained under identical air-oxidized conditions confirmed that the ratio of the symmetric form to the asymmetric form (that is, Fe2:Fe2′ in Supplementary Fig. 2) remained between 3:2 and 2:3. Full occupation of the asymmetric form has been observed only when an efficient oxidant such as ferricyanide is used. The symmetric form was recovered on reduction with H₂ or titanium(III) citrate, or by excessive exposure to X-rays on the air-oxidized crystals. A redox-dependent conformational

¹Department of Life Science, Graduate School of Life Science, University of Hyogo, 3-2-1 Koto, Kamigori-cho, Ako-gun, Hyogo 678-1297, Japan. ²RIKEN SPring-8 Center, 1-1-1 Koto, Sayo-gun, Sayo-cho, Hyogo 679-5148, Japan. ³Department of Bioresource Science, College of Agriculture, Ibaraki University, 3-21-1 Chu-ou, Ami-machi, Inashiki-gun, Ibaraki 300-0393, Japan. ⁴Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Kawaguchi Center Building, 4-1-8 Honcho, Kawaguchi-shi, Saitama 332-0012, Japan. †Present address: Department of Chemistry and Biochemistry, Graduate School of Engineering, Kyushu University, 744 Moto-oka, Nishi-ku, Fukuoka 819-0395, Japan.

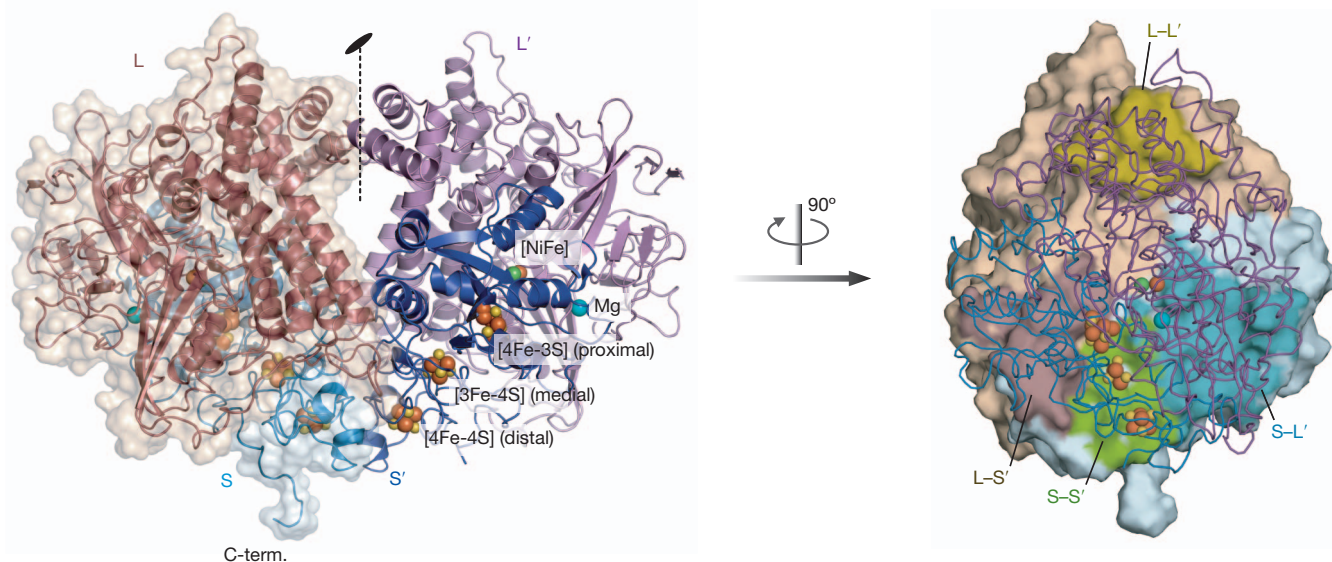


Figure 1 | Overall structure of MBH heterotetramer. The three structures in different oxidation conditions (Supplementary Table 2) are nearly identical to each other, therefore only the structure for the H_2 -reduced condition is shown. Two large subunits (labelled L and L') and two small subunits (S and S') are shown with different colours. One of the two heterodimers is represented by surface models with different colours. In the left panel (front view), a pseudo

change in an Fe-S cluster was also reported for the P-cluster in the nitrogenase from *Azotobacter vinelandii*¹⁵, a [4Fe-4S] cluster in the NADH-quinone oxidoreductase from *Thermus thermophilus*¹⁶, and hybrid cluster proteins from *Desulfovibrio vulgaris*¹⁷, of which the significance on their reaction mechanisms was pointed out. As for the coordination of the peptide amide nitrogen to the iron atom, this is the third reported example for metalloproteins^{15,18}.

The net charge of the [4Fe-3S]-6Cys in MBH established in this study is equal to that of the [4Fe-4S]-4Cys in the standard enzymes, assuming that the oxidation states of the iron atoms are identical. In addition, the numbers and positions of the charged residues around the proximal clusters of MBH in the symmetric form and the standard enzymes are similar except for the presence of two additional acidic residues in MBH (Fig. 3): Asp 51^S, the carboxy group of which locates ~ 12 Å from the edge of the proximal cluster, and Glu 73^L, which neutralizes nearby Arg 74^L. The fact that Glu 73^L is not strictly conserved among MBH (Supplementary Fig. 1b) indicates that the residue

dyad in the dimer of heterodimers is indicated; the C terminus of the small subunit (S) and the metal clusters in one of the heterodimers (L' and S') are labelled. The Fe-S clusters are termed proximal, medial and distal clusters with regard to their respective distance from the Ni-Fe active site. In the right panel (side view), the contact areas between the heterodimers are indicated with different colours to show which subunit is involved in the association.

has little effect on the electronic property of the proximal cluster. Taken together, the redox transition of $4+/3+$ is most probable for the proximal [4Fe-3S] cluster in the symmetric form during the H_2 -oxidation catalytic cycle; the oxidation states of the iron atoms are $2Fe^{2+}:2Fe^{3+}$ and $3Fe^{2+}:1Fe^{3+}$ as in standard [4Fe-4S] clusters. This conclusion does not conflict with the results of electron paramagnetic resonance studies on MBH from different bacteria^{19–21}—the midpoint potentials of the proximal cluster in the higher diamagnetic and the lower paramagnetic states of *R. eutropha* H16 MBH and *Aquifex aeolicus* MBH were reported to be -60 and $+87$ mV, respectively. Furthermore, a predicted superoxidized paramagnetic cluster for *R. eutropha* H16 and *A. aeolicus* should correspond to the asymmetric form observed in this study based on the measured midpoint potentials of $+160$ and $+232$ mV, respectively^{19,22}. The superoxidation of the proximal cluster to $[4Fe-3S]^{5+}$ should require the conformational change and concomitant donation of an additional negative charge described above. Previously proposed two single-electron transitions

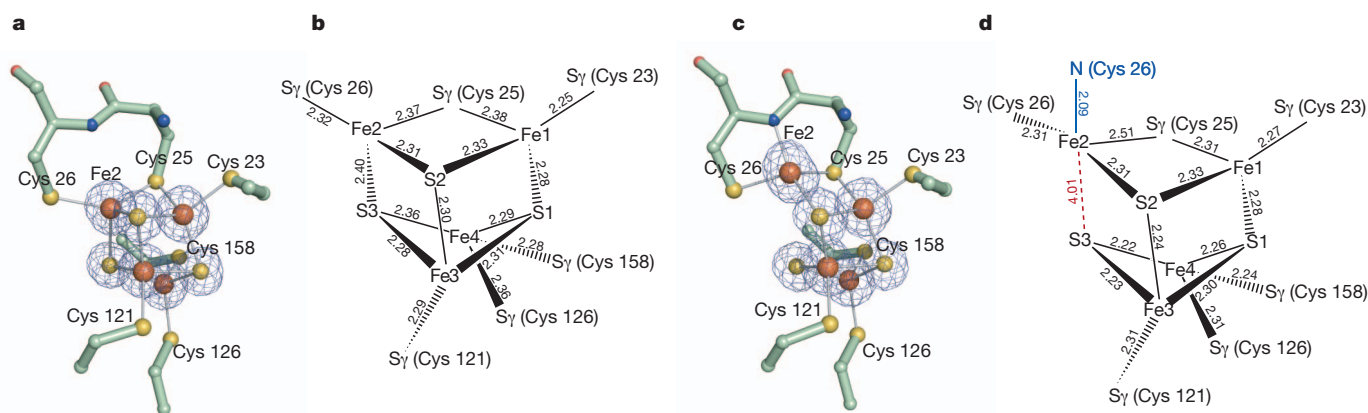


Figure 2 | Structural comparison of the proximal cluster in different redox conditions. a, b, H_2 -reduced condition. c, d, Ferricyanide-oxidized condition. Meshes represent $F_o - F_c$ maps contoured at 3.0σ where non-protein atoms were omitted in the calculation of F_c . The numbers shown in b and d represent

the bond distances (Å). The blue line indicates a bond that is newly formed by oxidation; the dashed red line indicates a bond that is found in the reduced form but is broken in the oxidized form.

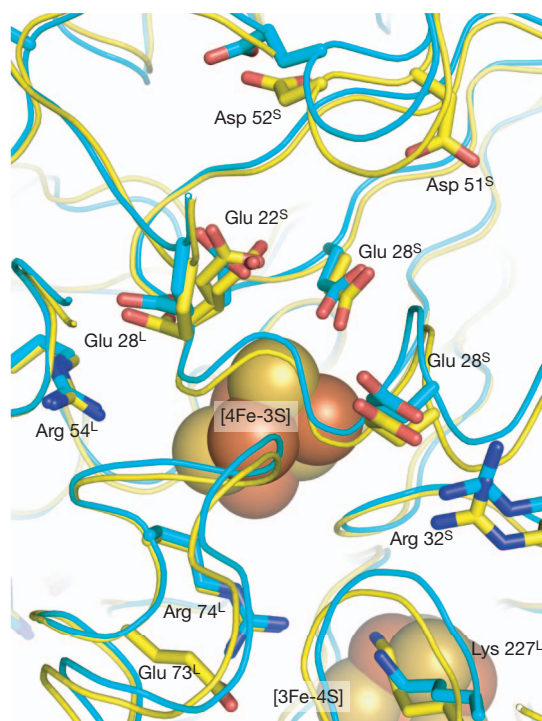


Figure 3 | A comparison of charged residues around the proximal Fe-S cluster in MBH and the standard [NiFe]-hydrogenase from *D. gigas*. Carbon atoms of MBH and the standard enzyme are shown in yellow and cyan, respectively. All charged residues shown are conserved between two hydrogenases, except Asp 51^S and Glu 73^L, which are only found in MBH.

of the proximal cluster in MBH (that is, $[4\text{Fe-3S}]^{5+/4+/3+}$)¹⁹ can thus be reasonably explained by the redox-dependent conformational change observed in this study.

After catalytic oxidation of H_2 at the [NiFe] active site buried deep in the protein, two protons have to be transferred to the solvent through a putative proton-transfer pathway(s). Although the details of the proton-transfer mechanism remain to be elucidated even for the standard enzymes, an initial proton-transfer step involving the protonation of Cys 530^L (residue numbering is for the standard enzyme from *Desulfovibrio gigas*) and the subsequent proton transfer to

Glu 18^L is widely accepted¹³ (Fig. 4a). This is based on the higher temperature factors of the two functional groups of Cys 530^L and Glu 18^L compared with those of the nearby atoms¹³, and on mutagenic studies using Glu25^LGln and Glu25^LAsp variants (Glu 25^L corresponds to Glu 18^L in the *D. gigas* enzyme) of *Desulfovibrio fructosovorans*²³. Two subsequent proton-transfer pathways that follow the protonation of Glu 18^L were proposed for the standard enzyme^{13,24}: one via the residues around the proximal Fe-S cluster and the other via the C-terminal histidine (His 536^L), the carboxy group of which coordinates the magnesium ion (Fig. 4a). The proton transfer from Cys 590^L (Cys 530^L in *D. gigas*) to Glu 28^L (Glu 18^L in *D. gigas*) is also likely to occur in MBH based on its structural similarity to the standard enzymes around the active site, but the alternative pathway to the C-terminal His 596^L is unlikely, where Met 52^L of MBH disturbs the array of water molecules that was identified in the standard enzymes (Fig. 4a, b). It seems that Met 52^L by itself does not cause the structural difference, but the nearby Leu 21^S affects the position of Met 52^L by steric hindrance. Actually, Leu 21^S is conserved only in MBH, whereas Met 52^L is found in [NiFeSe]-hydrogenases and a H_2 -sensing [NiFe]-hydrogenase (Supplementary Fig. 1a, b). On the other hand, the pathway via the residues around the proximal Fe-S cluster has been identified in MBH (Fig. 4b). Notably, this pathway has a connection to the amide nitrogen that coordinates Fe2 in the asymmetric form of the proximal Fe-S cluster, indicating that the observed redox-dependent structural change in the proximal cluster is directly coupled to the redox states and reactions at the Ni-Fe active site, and that one proton liberated from the amide-hydrogen can be transferred to the active site to reduce O_2 , as described below.

The standard enzyme is known to be inactivated reversibly to form two spectroscopically distinguishable states, Ni-A and Ni-B (ref. 25). Whereas Ni-A is preferentially produced by the oxidation of the active site with O_2 under electron-deficient conditions and requires considerable time for the reductive reactivation, Ni-B is formed in electron-rich conditions and is readily reactivated in a few seconds²⁶. It has been suggested that preventing the Ni-Fe cluster from being trapped in the inactive Ni-A state is crucial for MBH to function in the presence of O_2 (refs 4, 22, 27). Therefore, having a higher redox potential and a two-electron donation property in the proximal cluster is advantageous in preventing the formation of Ni-A, as previously pointed out^{19,22}. The structure of the Ni-Fe active site of MBH shows close similarity to those of the standard enzymes (Supplementary Fig. 3a, b)¹³, excluding the possibility that the O_2 tolerance of MBH

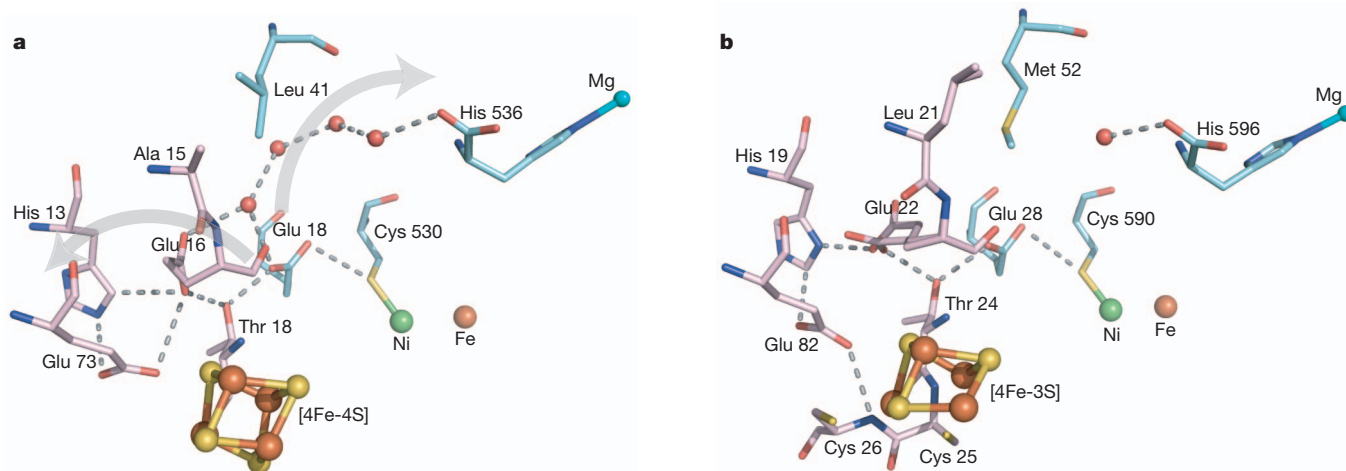


Figure 4 | Proposed proton-transfer pathways. **a**, The standard [NiFe]-hydrogenase from *D. gigas* (Protein Data Bank accession 1QY9). **b**, MBH in this study. The two proton-transfer pathways downstream of Glu 18^L (Glu 28^L in MBH) proposed in the standard enzymes are indicated with grey arrows in **a**. The carbon atoms of the large and small subunits are shown in cyan and pink,

respectively. The hydrogen bonds are represented by dotted grey lines. The red spheres are water molecules identified in the crystal structures. For His 19^S (His 13^S in *D. gigas*), the inversion of the imidazole ring is taken into account for the formation of hydrogen bonds. His 596^L (His 536^L in *D. gigas*) is the C-terminal residue of the large subunit.

is ascribable to the structural difference in the active site as suggested for [NiFeSe]-hydrogenase²⁸. Furthermore, our structural analysis has confirmed that the physical exclusion of O₂ by the H₂ channel, proposed for H₂-sensing [NiFe]-hydrogenases^{29,30}, is not the case for MBH, because the size and shape of the channels of MBH and the standard enzyme of *D. vulgaris* MF showed no particular difference that would affect the diffusion of O₂ (Supplementary Fig. 4a, b). Finally, a recent study²² using an MBH mutant that lacks the additional two cysteine residues in the proximal cluster has shown that the cluster is important for O₂ tolerance both *in vivo* and *in vitro*. Our results provide a structural basis that underlies the unprecedented function of the [4Fe-3S] cluster in the O₂ tolerance of MBH, which acquires a superoxidized state to supply two electrons and one proton for the reduction of O₂ by using a redox-dependent conformational change.

METHODS SUMMARY

MBH crystals were prepared anaerobically by mixing 2 µl of the protein solution consisting of 27.5 mg ml⁻¹ MBH, 10 mM Tris-HCl (pH 7.4) and 2 mM DTT with 2 µl of the reservoir solution including 100 mM PIPES (pH 6.5), 15% PEG3350, 300 mM Li₂SO₄ and 5 mM DTT. To prepare the H₂-reduced crystals, fresh crystals were soaked in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 1 mM methyl viologen in a screw-capped vial with a silicon septum, and the crystals were reduced under 100 kPa of H₂ for 2 h at 313 K. The ferricyanide-oxidized crystals were produced by soaking the fresh crystals in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 5 mM K₃[Fe(CN)₆] for 1 h at room temperature (between 293 and 298 K) inside a glove box. The air-oxidized crystals were prepared by soaking the fresh crystals in the same buffer as for H₂-reduced crystals but without methyl viologen for 1 h at room temperature outside the glove box. All crystals were cooled with liquid nitrogen inside the glove box with the exception of the air-oxidized crystals.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 April; accepted 23 August 2011.

Published online 16 October 2011.

- Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
- Vincent, K. A. *et al.* Electrocatalytic hydrogen oxidation by an enzyme at high carbon monoxide or oxygen levels. *Proc. Natl Acad. Sci. USA* **102**, 16951–16954 (2005).
- Luo, X., Brugna, M., Tron-Infossi, P., Giudici-Orticoni, M. T. & Lojou, E. Immobilization of the hyperthermophilic hydrogenase from *Aquifex aeolicus* bacterium onto gold and carbon nanotube electrodes for efficient H₂ oxidation. *J. Biol. Inorg. Chem.* **14**, 1275–1288 (2009).
- Lukey, M. J. *et al.* How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J. Biol. Chem.* **285**, 3928–3938 (2010).
- De Lacey, A. L., Fernandez, V. M., Rousset, M. & Cammack, R. Activation and inactivation of hydrogenase function and the catalytic cycle: spectroelectrochemical studies. *Chem. Rev.* **107**, 4304–4330 (2007).
- Brugna-Guiral, M. *et al.* [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles* **7**, 145–157 (2003).
- Burgdorf, T. *et al.* [NiFe]-hydrogenases of *Ralstonia eutropha* H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation. *J. Mol. Microbiol. Biotechnol.* **10**, 181–196 (2005).
- Yoon, K.-S., Fukuda, K., Fujisawa, K. & Nishihara, H. Purification and characterization of a highly thermostable, oxygen-resistant, respiratory [NiFe]-hydrogenase from a marine, aerobic hydrogen-oxidizing bacterium *Hydrogenovibrio marinus*. *Int. J. Hydrogen Energy* **36**, 7081–7088 (2011).
- Volbeda, A. *et al.* Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).
- Higuchi, Y., Yagi, T. & Yasuoka, N. Unusual ligand structure in Ni-Fe active center and an additional Mg site in hydrogenase revealed by high resolution X-ray structure analysis. *Structure* **5**, 1671–1680 (1997).
- Montet, Y. *et al.* Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nature Struct. Biol.* **4**, 523–526 (1997).
- Matias, P. M. *et al.* [NiFe] hydrogenase from *Desulfovibrio desulfuricans* ATCC 27774: gene sequencing, three-dimensional structure determination and refinement at 1.8 Å and modelling studies of its interaction with the tetraheme cytochrome c₃. *J. Biol. Inorg. Chem.* **6**, 63–81 (2001).

- Fontecilla-Camps, J. C., Volbeda, A., Cavazza, C. & Nicolet, Y. Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chem. Rev.* **107**, 4273–4303 (2007).
- Saggu, M. *et al.* Spectroscopic insights into the oxygen-tolerant membrane-associated [NiFe] hydrogenase of *Ralstonia eutropha* H16. *J. Biol. Chem.* **284**, 16264–16276 (2009).
- Peters, J. W. *et al.* Redox-dependent structural changes in the nitrogenase P-cluster. *Biochemistry* **36**, 1181–1187 (1997).
- Berrisford, J. M. & Sazanov, L. A. Structural basis for the mechanism of respiratory complex I. *J. Biol. Chem.* **284**, 29773–29783 (2009).
- Aragão, D., Mitchell, E. P., Frazao, C. F., Carrondo, M. A. & Lindley, P. F. Structural and functional relationships in the hybrid cluster protein family: structure of the anaerobically purified hybrid cluster protein from *Desulfovibrio vulgaris* at 1.35 Å resolution. *Acta Crystallogr. D* **64**, 665–674 (2008).
- Huang, W. *et al.* Crystal structure of nitrile hydratase reveals a novel iron centre in a novel fold. *Structure* **5**, 691–699 (1997).
- Pandelia, M. E. *et al.* Characterization of a unique [FeS] cluster in the electron transfer chain of the oxygen tolerant [NiFe] hydrogenase from *Aquifex aeolicus*. *Proc. Natl Acad. Sci. USA* **108**, 6097–6102 (2011).
- Schneider, K., Patil, D. S. & Cammack, R. ESR properties of membrane-bound hydrogenase from aerobic hydrogen bacteria. *Biochim. Biophys. Acta* **748**, 353–361 (1983).
- Knüttel, K. *et al.* Redox properties of the metal centres in the membrane-bound hydrogenase from *Alcaligenes eutrophus* CH34. *Bull. Pol. Acad. Sci. Chem.* **42**, 495–511 (1994).
- Goris, T. *et al.* A unique iron-sulfur cluster is crucial for oxygen tolerance of a [NiFe]-hydrogenase. *Nature Chem. Biol.* **7**, 310–318 (2011).
- Dementin, S. *et al.* A glutamate is the essential proton transfer gate during the catalytic cycle of the [NiFe] hydrogenase. *J. Biol. Chem.* **279**, 10508–10513 (2004).
- Teixeira, V. H., Soares, C. M. & Baptista, A. M. Proton pathways in a [NiFe]-hydrogenase: A theoretical study. *Proteins* **70**, 1010–1022 (2008).
- Lubitz, W., Reijerse, E. & van Gestel, M. [NiFe] and [FeFe] hydrogenases studied by advanced magnetic resonance techniques. *Chem. Rev.* **107**, 4331–4365 (2007).
- Fernandez, V. M., Hatchikian, E. C., Patil, D. S. & Cammack, R. ESR-detectable nickel and iron-sulphur centres in relation to the reversible activation of *Desulfovibrio gigas* hydrogenase. *Biochim. Biophys. Acta* **883**, 145–154 (1986).
- Cracknell, J. A., Wait, A. F., Lenz, O., Friedrich, B. & Armstrong, F. A. A kinetic and thermodynamic understanding of O₂ tolerance in [NiFe]-hydrogenases. *Proc. Natl Acad. Sci. USA* **106**, 20681–20686 (2009).
- Marques, M. C., Coelho, R., De Lacey, A. L., Pereira, I. A. & Matias, P. M. The three-dimensional structure of [NiFeSe] hydrogenase from *Desulfovibrio vulgaris* Hildenborough: a hydrogenase without a bridging ligand in the active site in its oxidised, “as-isolated” state. *J. Mol. Biol.* **396**, 893–907 (2010).
- Buhrke, T., Lenz, O., Krauss, N. & Friedrich, B. Oxygen tolerance of the H₂-sensing [NiFe] hydrogenase from *Ralstonia eutropha* H16 is based on limited access of oxygen to the active site. *J. Biol. Chem.* **280**, 23791–23796 (2005).
- Duché, O., Elsen, S., Cournaud, L. & Colbeau, A. Enlarging the gas access channel to the active site renders the regulatory hydrogenase HupUV of *Rhodobacter capsulatus* O₂ sensitive without affecting its transducing activity. *FEBS J.* **272**, 3899–3908 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Hagiya for technical assistance at Ibaraki University. The synchrotron radiation experiments were performed at the BL41XU (proposal no. 2010A1223) and BL44XU (proposal no. 2010A/B6520) with the approval of JASRI. The CCD detector MX225-HE (Rayonix) at BL44XU was financially supported by Academia Sinica and National Synchrotron Radiation Research Center (Taiwan). This work was supported by a grant-in-aid for Scientific Research from MEXT (20051022 (Y.S.), 22770111 (Y.S.), 18GS0207 (Y.H.)), grant-in-aid for Scientific Research from JSPS (20580094 (H.N.), 22370061 (Y.H.) and 22657031 (Y.H.)), grant-in-aid for research and education from University of Hyogo (Y.S.), and grant-in-aid for young scientists from Hyogo Science and Technology Association (Y.S.). This work was also partially supported by the GCOE Program (Y.H.), the Japanese Aerospace Exploration Agency Project (Y.H.), and the basic research programs CREST type, ‘Development of the Foundation for Nano-Interface Technology’ from JST, Japan (Y.H.).

Author Contributions K.-S.Y. and H.N. performed bacterial culture and protein purification under management by H.N. Y.S. performed crystallization, X-ray data collection and structure determination. Y.S. and Y.H. prepared the manuscript with contributions from all authors. The overall project management was done by Y.H.

Author Information Atomic coordinates and structure factors for the reported structures have been deposited in the Protein Data Bank with the accession codes 3AYX, 3AYY and 3AYZ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.H. (hig@sci.u-hyogo.ac.jp).

METHODS

Production and preparation of crystals. The procedures for bacterial cultivation, purification and crystallization of MBH, and the preliminary crystallographic experiment have been described previously³¹. Crystallization was performed inside an anaerobic glove box (Model F1024, Thermo Fisher Scientific) by the sitting-drop vapour diffusion method at room temperature. The inner solution was prepared by mixing 2 μ l of protein solution consisting of 27.5 mg ml⁻¹ MBH, 10 mM Tris-HCl (pH 7.4), and 2 mM DTT with 2 μ l of reservoir solution including 100 mM PIPES (pH 6.5), 15% PEG3350, 300 mM Li₂SO₄ and 5 mM DTT. Dark-brown crystals grew to maximum size ($\sim 0.2 \times 0.2 \times 0.1$ mm) within a few days. To prepare the H₂-reduced crystals, fresh crystals were soaked in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 1 mM methyl viologen in a screw-capped vial with silicon septum. The reduction of crystals was performed under 1 bar of H₂ for 2 h at 313 K, and was optically confirmed by the blue colour of reduced methyl viologen. The ferricyanide-oxidized crystals were produced by soaking the fresh crystals in buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄, 20% glycerol and 5 mM K₃[Fe(CN)₆] for 1 h at room temperature inside the glove box. The air-oxidized crystals were prepared by soaking the fresh crystals in air-saturated buffer containing 100 mM PIPES (pH 6.5), 20% PEG3350, 300 mM Li₂SO₄ and 20% glycerol for 1 h at room temperature outside the glove box. All crystals were cooled with liquid nitrogen inside the glove box with the exception of the air-oxidized crystals.

X-ray diffraction data collection and structural refinement. The X-ray diffraction data were collected at beamline BL41XU and BL44XU of SPring-8 (Hyogo, Japan) using Beamline Schedule Software (BSS)³². The crystals were maintained at 90 K using a gaseous nitrogen stream during data collection. The diffraction data were integrated and scaled with the program HKL2000 (ref. 33). The structure was

determined by the molecular replacement method using the program MolRep from the CCP4 suite³⁴ with atomic coordinates of the enzyme from *D. vulgaris* MF (Protein Data Bank accession 1WUJ) as a search model. Refinement with isotropic B-factors for individual atoms was performed with the program Refmac5 (ref. 35), and subsequent refinement with anisotropic B-factors and alternative conformations were carried out with the program SHELX-97 (ref. 36) without merging diffractions in Bijvoet pairs. For tentatively assigned oxygen atoms around cysteine thiolates in the [NiFe] active centre and Fe2 in the proximal Fe-S cluster, occupancy and B-factor values were alternately refined. In the final step of each refinement, hydrogen atoms were included. Visual inspection and manual correction of models were performed with the program COOT³⁷. Data collection and refinement statistics are summarized in Supplementary Table 2. Figures were prepared with using the program PyMol (DeLano Scientific).

31. Shomura, Y., Hagiya, K., Yoon, K.-S., Nishihara, H. & Higuchi, Y. Crystallization and preliminary X-ray diffraction analysis of membrane-bound respiratory [NiFe] hydrogenase from *Hydrogenovibrio marinus*. *Acta Crystallogr. F* **67**, 827–829 (2011).
32. Ueno, G., Kanda, H., Kumasaka, T. & Yamamoto, M. Beamline Scheduling Software: administration software for automatic operation of the RIKEN structural genomics beamlines at SPring-8. *J. Synchrotron Radiat.* **12**, 380–384 (2005).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
34. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
36. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

CAREERS

TURNING POINT From medicine in China to cancer research in the United States **p.258**

ADVICE FORUM Get expert input on scientific career issues go.nature.com/lm1x4t

NATUREJOBS For the latest career listings and advice www.naturejobs.com



COLUMN

Part-time balance

Flexible academic positions help women to juggle work and family. **Kate O'Brien** and **Karen Hapgood** explain how to avoid the 'female ghetto' when working part time.

When scientists become mothers, many seek part-time positions to maintain their careers while still spending significant time with their families. But they might not anticipate the downside: part-timers can lose research time and seniority. In academia, they risk being funnelled into teaching positions that can exclude them from research. But there are ways around this.

Part-timers are generally at a disadvantage when competing with full-time colleagues. Full-timers' higher productivity leads to more funding and attracts top-quality collaborators, postdocs and graduate students, so part-time researchers, who are often women, can fall behind. Teaching and administration further reduce the part-timer's research time. And yet they are frequently assessed on the same scale as their full-time colleagues, using metrics that do not account for either their part-time status or the longer time they spend in 'early-career'

phase, when output is understandably low. An absence of role models exacerbates this; most or all other academics in the department, particularly senior ones, will be full-time tenure-track. This creates a discouraging situation in which capable women working at a high level may still be judged poorly. There is also a risk of a 'female ghetto' — a tier of women with fewer opportunities for promotion and job security.

A PATH TO SUCCESS

The part-time model is much more viable in teaching-focused roles, in which evaluation is often based on current performance rather than on accumulated output. Also, teaching does not have the same minimum participation rate as research, and allows ready flexibility. Part-time staff naturally drift towards teaching, or can be pushed there by management — they are often allocated disproportionately high teaching loads because of their

reduced research productivity. Consequently, many female academics focus on teaching at the expense of research. Teaching contracts can be a workable, flexible and fun way to remain engaged with science while your children grow, but it is difficult to re-enter research from teaching later.

Yet it is not impossible to pursue a part-time career in scientific research while devoting time and creative energy to raising children. Choose a role that lets you maintain and build your skills, and define success on your own terms. Developing a research group and teaching in a tenure-track role will be difficult part time unless you are already quite advanced in your career; a contract role in a strong research group might bring less prestige and security, but will let you develop your expertise and build a track record in preparation for a successful tenured position when your children are older.

Choose an organization that is friendly to ►

BELLE MELLOR

► part-time workers. You will need a manager who can appreciate your contribution and nurture your development. Find out about the presence and status of other part-time scientists; schemes that support career re-entry; and whether there are limitations on part-time staff supervising PhD students or applying for funding. A good manager will recognize that your worth is unlikely to be directly reflected in metrics that have been developed for and by full-time academics. Be sure that your institution, directors and manager will allocate part-time teaching and administration loads fairly, and are willing to adjust metrics and milestones to account for career stage, part-time status and other duties (such as teaching).

If you are in a tenured position, take the initiative and find solutions that will work for you and your department. Negotiate for all your teaching duties to be in one semester, so you can dedicate the other to continuous research. Collaborate with or join a research team in which your expertise is valued, your part-time status is accepted and you can work with established researchers. This may require compromise and concentration on one key research area.

Recognize where part-time work provides an advantage over more conventional employment. A part-time postdoc, research fellow or contract researcher can adapt to workloads that vary with the cycles of grant funding, and provide continuity and management in a lab group. At the same time, the part-timer can maintain and build their own expertise so that they can embark on more ambitious projects if they return to full-time work after a few years, or even decades.

You need to be brave to undertake an unusual career path. You may be unsuccessful according to the metrics used by administrators, and your performance may be judged poorly against that of full-time scientists. Maintaining confidence in your abilities will be difficult under these conditions. Wise mentoring, and acquaintance with others who have worked part-time or follow non-traditional career paths, can help you to negotiate the system.

You must be patient with your ambitions, and with your managers and colleagues as they adapt to working with part-time researchers. Above all, enjoy your time with your children. Your full-time colleagues will be promoted ahead of you, but there will be time to focus more on your career once your children have grown. ■

Kate O'Brien is a part-time lecturer in chemical and environmental engineering at the University of Queensland in Brisbane, Australia. **Karen Hapgood** is a full-time associate professor in chemical engineering at Monash University in Melbourne, Australia.

TURNING POINT

Qin Liu

In July, Qin Liu began her tenure as an associate professor of biostatistics at the Wistar Institute in Philadelphia, Pennsylvania. She explains how she has combined a medical degree with graduate research in epidemiology and biostatistics to focus on cancer research.

You set out to be a clinical doctor. What sparked your interest in biostatistics?

I expected to become a clinical doctor when I completed my MD at Shanxi Medical University in Taiyuan, China. I got a job offer at a hospital in Changzhi, my parents' current residence. Then, during the last year of medical school, I did an internship at the Center for Disease Control and Prevention in Shanxi Province. I wanted to use population-based data to get useful public-health information. I thought I would do a graduate programme in epidemiology, but the principal investigator wanted a man. He felt it wasn't safe for a woman to travel alone collecting data in the countryside. Biostatistics was similar, without the travel, and a good fit. I did a survival analysis of people with lung cancer.

How did you end up in the United States?

After completing my PhD at Shanghai Medical University and working there as a lecturer, I was chosen, from 30 applicants, for the one opening in a biostatistics postdoctoral training programme at the Cancer Center of the University of Massachusetts (UMass) in Worcester.

What was a pivotal point in your early research?

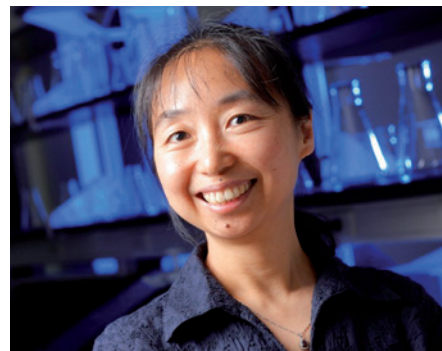
I found that women who get pregnant have an increased risk of developing breast cancer for several years after pregnancy — but later in life, their risk decreases. My team and I interpreted this as evidence that the risk might be associated with hormonal changes. That was a big paper for me. Several years later, using a similar statistical model applied to the same Swedish population data, we found the opposite trend for ovarian cancer. We need more data to understand this phenomenon.

Why did you then pursue a master's degree in epidemiology?

My mentor was an epidemiologist, and I felt that I needed more knowledge for population-based studies. At the time, UMass employees got a 50% reimbursement for work-related degrees, so I thought, 'Why not?' I was also curious as to how US university teachers teach.

What are the biggest differences between Chinese and US teachers?

In China, teachers seldom ask students'



WISTAR

opinions or discuss topics with them. When I taught, I talked for the entire 45 minutes of class. But here, they use half their time to lead a discussion and ask students questions. It makes the students feel active. I was nervous at first as a student, because my English was not good and I couldn't organize my thoughts to reply. But I learned to focus my energy to be ready when the teacher called on me.

Do you think you will eventually return to China?

Several people have contacted me, and I know there is a lot of funding from the Chinese government to attract good researchers back to help develop public-health projects. But so far, I haven't thought about going back.

How have funding ups and downs affected your career?

I worked at the cancer centre at UMass for seven years, until our funding ran out. Almost everyone in our group had to look for another position. It was very hard. This is a challenge in our field: there are few independent grant-ing sources, so biostatisticians have to either collaborate with other people to analyse their data or work in a group that provides statistical services to faculty members.

I was lucky that the UMass Biostatistical Research Group, which provides statistical services to the entire department of medicine, had an opening and chose me. I worked with all different types of researchers, conducting everything from clinical to policy studies. But I realized I really wanted to return to cancer research when I saw this opportunity in the oncogenesis programme at Wistar. Luckily, the hiring committee saw me as a potential bridge between the basic science and the clinical research being done there. I hope the position will afford me more research independence. ■

INTERVIEW BY VIRGINIA GEWIN

► part-time workers. You will need a manager who can appreciate your contribution and nurture your development. Find out about the presence and status of other part-time scientists; schemes that support career re-entry; and whether there are limitations on part-time staff supervising PhD students or applying for funding. A good manager will recognize that your worth is unlikely to be directly reflected in metrics that have been developed for and by full-time academics. Be sure that your institution, directors and manager will allocate part-time teaching and administration loads fairly, and are willing to adjust metrics and milestones to account for career stage, part-time status and other duties (such as teaching).

If you are in a tenured position, take the initiative and find solutions that will work for you and your department. Negotiate for all your teaching duties to be in one semester, so you can dedicate the other to continuous research. Collaborate with or join a research team in which your expertise is valued, your part-time status is accepted and you can work with established researchers. This may require compromise and concentration on one key research area.

Recognize where part-time work provides an advantage over more conventional employment. A part-time postdoc, research fellow or contract researcher can adapt to workloads that vary with the cycles of grant funding, and provide continuity and management in a lab group. At the same time, the part-timer can maintain and build their own expertise so that they can embark on more ambitious projects if they return to full-time work after a few years, or even decades.

You need to be brave to undertake an unusual career path. You may be unsuccessful according to the metrics used by administrators, and your performance may be judged poorly against that of full-time scientists. Maintaining confidence in your abilities will be difficult under these conditions. Wise mentoring, and acquaintance with others who have worked part-time or follow non-traditional career paths, can help you to negotiate the system.

You must be patient with your ambitions, and with your managers and colleagues as they adapt to working with part-time researchers. Above all, enjoy your time with your children. Your full-time colleagues will be promoted ahead of you, but there will be time to focus more on your career once your children have grown. ■

Kate O'Brien is a part-time lecturer in chemical and environmental engineering at the University of Queensland in Brisbane, Australia. **Karen Hapgood** is a full-time associate professor in chemical engineering at Monash University in Melbourne, Australia.

TURNING POINT

Qin Liu

In July, Qin Liu began her tenure as an associate professor of biostatistics at the Wistar Institute in Philadelphia, Pennsylvania. She explains how she has combined a medical degree with graduate research in epidemiology and biostatistics to focus on cancer research.

You set out to be a clinical doctor. What sparked your interest in biostatistics?

I expected to become a clinical doctor when I completed my MD at Shanxi Medical University in Taiyuan, China. I got a job offer at a hospital in Changzhi, my parents' current residence. Then, during the last year of medical school, I did an internship at the Center for Disease Control and Prevention in Shanxi Province. I wanted to use population-based data to get useful public-health information. I thought I would do a graduate programme in epidemiology, but the principal investigator wanted a man. He felt it wasn't safe for a woman to travel alone collecting data in the countryside. Biostatistics was similar, without the travel, and a good fit. I did a survival analysis of people with lung cancer.

How did you end up in the United States?

After completing my PhD at Shanghai Medical University and working there as a lecturer, I was chosen, from 30 applicants, for the one opening in a biostatistics postdoctoral training programme at the Cancer Center of the University of Massachusetts (UMass) in Worcester.

What was a pivotal point in your early research?

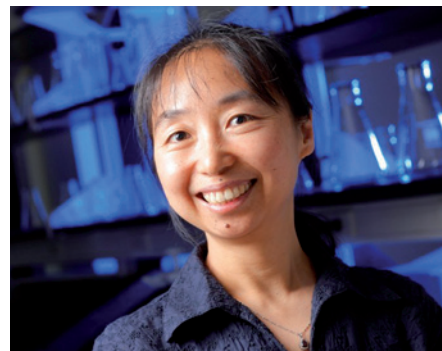
I found that women who get pregnant have an increased risk of developing breast cancer for several years after pregnancy — but later in life, their risk decreases. My team and I interpreted this as evidence that the risk might be associated with hormonal changes. That was a big paper for me. Several years later, using a similar statistical model applied to the same Swedish population data, we found the opposite trend for ovarian cancer. We need more data to understand this phenomenon.

Why did you then pursue a master's degree in epidemiology?

My mentor was an epidemiologist, and I felt that I needed more knowledge for population-based studies. At the time, UMass employees got a 50% reimbursement for work-related degrees, so I thought, 'Why not?' I was also curious as to how US university teachers teach.

What are the biggest differences between Chinese and US teachers?

In China, teachers seldom ask students'



WISTAR

opinions or discuss topics with them. When I taught, I talked for the entire 45 minutes of class. But here, they use half their time to lead a discussion and ask students questions. It makes the students feel active. I was nervous at first as a student, because my English was not good and I couldn't organize my thoughts to reply. But I learned to focus my energy to be ready when the teacher called on me.

Do you think you will eventually return to China?

Several people have contacted me, and I know there is a lot of funding from the Chinese government to attract good researchers back to help develop public-health projects. But so far, I haven't thought about going back.

How have funding ups and downs affected your career?

I worked at the cancer centre at UMass for seven years, until our funding ran out. Almost everyone in our group had to look for another position. It was very hard. This is a challenge in our field: there are few independent grant-ing sources, so biostatisticians have to either collaborate with other people to analyse their data or work in a group that provides statistical services to faculty members.

I was lucky that the UMass Biostatistical Research Group, which provides statistical services to the entire department of medicine, had an opening and chose me. I worked with all different types of researchers, conducting everything from clinical to policy studies. But I realized I really wanted to return to cancer research when I saw this opportunity in the oncogenesis programme at Wistar. Luckily, the hiring committee saw me as a potential bridge between the basic science and the clinical research being done there. I hope the position will afford me more research independence. ■

INTERVIEW BY VIRGINIA GEWIN

REMEMBER YUGOSLAVIA?

A crash course in business skills.

BY GARETH D. JONES

The board of directors sat across the table from Blake, an array of dark suits and expectant faces, waiting for him to speak. He had no idea what he was doing there.

I'm a salesman, darn it! I always know what to say. What worried Blake the most was that he couldn't remember standing up to address the board. He could ad-lib with the best of them, if he had a clue what the subject was.

"So, are you going to tell us about your trip to Yugoslavia?" the chairman asked, frowning.

Yugoslavia? Have I been there?

"Let me give you the background," Maeve stood up, auburn curls bouncing energetically, and smoothly took the floor. Blake sank into his chair gratefully.

"As you know, the international sports expo was the reunited republic's first big chance to show the rest of Europe what it had to offer..."

As Maeve talked, flashes of memory began to come to Blake: the flight from Stansted, the huge weather dome housing the expo, the sports facilities and exhibitions, the tennis coach...

"Now, *that's* impressive," Maeve said as they paused to watch the tennis match. For a change they weren't selling anything. Sport-Tech UK had sent them to check out the competition, look for market opportunities, partnerships, something to give them the edge.

"Could be a ringer," Blake said. The tennis pro — a well-known Belgian champion — was facing tough competition from his opponent, supposedly a passing member of the public who had never played before. The woman wore a glossy black headband fixed tightly to her skull. Blake read the promo leaflet that the saleswoman who was dressed as a ball girl gave to him.

"Apparently the headband transmits the coach's skills," he said, intrigued. The coach sat in a motorized wheelchair at the far side of the tennis court, a matching headband affixed to his skull.

"That's just showing off," Maeve snorted. "They didn't need to use a paraplegic coach."

Blake shrugged, and watched the match for a few more minutes until Maeve wandered off elsewhere. It was impressive technology, but more high-end than they were looking for.

There was a blank in his memory.



"Oh, go on, I'll give it a go," Blake said, allowing the ball girl to slide the headband snugly in position.

"There," she said with a pleasant Flemish accent. "It's not just for tennis of course." She showed him three tennis balls. "Do you juggle?"

"No."

"Well I do." She tapped her own headband and passed him the balls, one at a time. "Give it a go."

Blake weighed the tennis balls for a few seconds, then tossed the first in the air, followed by the second and the third. He was juggling! Then he dropped them.

"You tried to think about what you were doing." She smiled. "It's best to allow your subconscious to accept my direction." She picked the balls up. "Your mind can be trained to accept the skills from your coach."

"Did you hear?" Maeve asked, slipping into the seat next to Blake at the bar. The place was warm and noisy, packed with off-duty salesmen and sports enthusiasts.

Blake sipped his cider. "Hear what, particularly?"

"About the expo official, caught supplying financial information to an industrial spy?"

"No?" Blake's interest was piqued. Something other than sales to think about...

"She denies it of course, says she doesn't remember the meeting, even though it's caught on security footage."

Blake downed the remainder of his drink.

"Maybe I won't remember meeting up with you either," he said, and called for another pint. He was still irritated at

the way she'd taken over his chat with a pair of well-tanned sales execs elsewhere in the expo. According to their sales patter, their tablets stimulated melatonin levels to produce a natural tan in only an hour. Exactly the kind of cheap, mass-market product they were looking for.

The chairman appraised Blake over steepled fingers. "So Maeve has come back with a deal that makes us sole distributors of EasiTan in the UK, worth, what? Six million pounds?" She looked at Maeve.

"About six point two."

"Good. Excellent. And what did you get out of your little jaunt, Blake?"

There was an uncomfortable silence. Uncomfortable for Blake at least.

"I have to say that Blake was very good at sounding out the possibilities," Maeve offered.

Blake wasn't sure it helped.

"Next year's sports expo is in Belgium," the chairman said. "Let's hope you're still with us, shall we?"

Blake fiddled with the coins in his pocket as they took the lift down two floors, worried over his bizarrely fractured memories. "That didn't go so well," he said.

Maeve patted his shoulder sympathetically. "Next time keep your eye on the ball, not the ball girl."

"Ha!" He stuck out his tongue as she stepped out onto the sales office floor.

"I can't believe the mess Yugoslavia's in," she said as they walked the open-plan area to their desks.

"It is?"

"Weren't you paying any attention while we were there? The financial scandal has brought the government down." They sat at neighbouring desks and Blake leaned back in his chair as Maeve spoke. "The expo should've been worth a fortune to the economy, but it looks like Yugoslavia's first year together could be its last."

That gave Blake an idea.

"Are you listening?"

"Sorry, got something I need to do." He left his paper-strewn desk and headed back to the lift, feeling the urge to look up some company financial information. He pulled three pound coins from his pocket and juggled them as he walked. ■

Gareth D. Jones is an environmental scientist from the United Kingdom who also writes stories and drinks lots of tea.

➔ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

Antarctic accumulation seasonality

ARISING FROM T. Laepple, M. Werner & G. Lohmann *Nature* **471**, 91–94 (2011)

The resemblance of the orbitally filtered isotope signal from the past 340 kyr in Antarctic ice cores to Northern Hemisphere summer insolation intensity has been used to suggest that the northern hemisphere may drive orbital-scale global climate changes¹. A recent Letter² by Laepple *et al.* suggests that, contrary to this interpretation, this semblance may instead be explained by weighting the orbitally controlled Antarctic seasonal insolation cycle with a static (present-day) estimate of the seasonal cycle of accumulation. We suggest, however, that both time variability in accumulation seasonality and alternative stable seasonality can markedly alter the weighted insolation signal. This indicates that, if the last 340 kyr of Antarctic accumulation has not always looked like the estimate of precipitation and accumulation seasonality made by Laepple *et al.*², this particular accumulation weighting explanation of the Antarctic orbital-scale isotopic signal might not be robust.

Laepple *et al.*² present a range of present-day accumulation and precipitation estimates from a variety of sites across East Antarctica (figure 2a in ref. 2). A broad agreement on the pattern of seasonality suggests a

reasonable understanding of the present-day Antarctic precipitation seasonality. Laepple *et al.* use a geographical mean of these present-day records and an assumption of relatively stable seasonality of precipitation over the past 340 kyr in the calculations. However, Antarctic ice-core evidence suggests that during colder climatic periods, which comprise the majority of the record (Fig. 1b), East Antarctic Plateau precipitation was reduced; during the coldest glacial periods it was probably about half the present-day value³. This observational evidence suggests that changes in precipitation seasonality are possible.

Estimates of East Antarctic seasonality across different climates can be obtained using general circulation models. Estimates of Dome C accumulation seasonality from present-day, glacial maximum and warm climate simulations⁴ are shown in Fig. 1a. The simulation of present-day accumulation amount is accurate⁵. Relative to the present-day simulation, the cold climate simulation accumulation is approximately halved. This matches ice-core evidence³. The simulated present seasonal accumulation cycle at this site is within the range of the observational uncertainty².

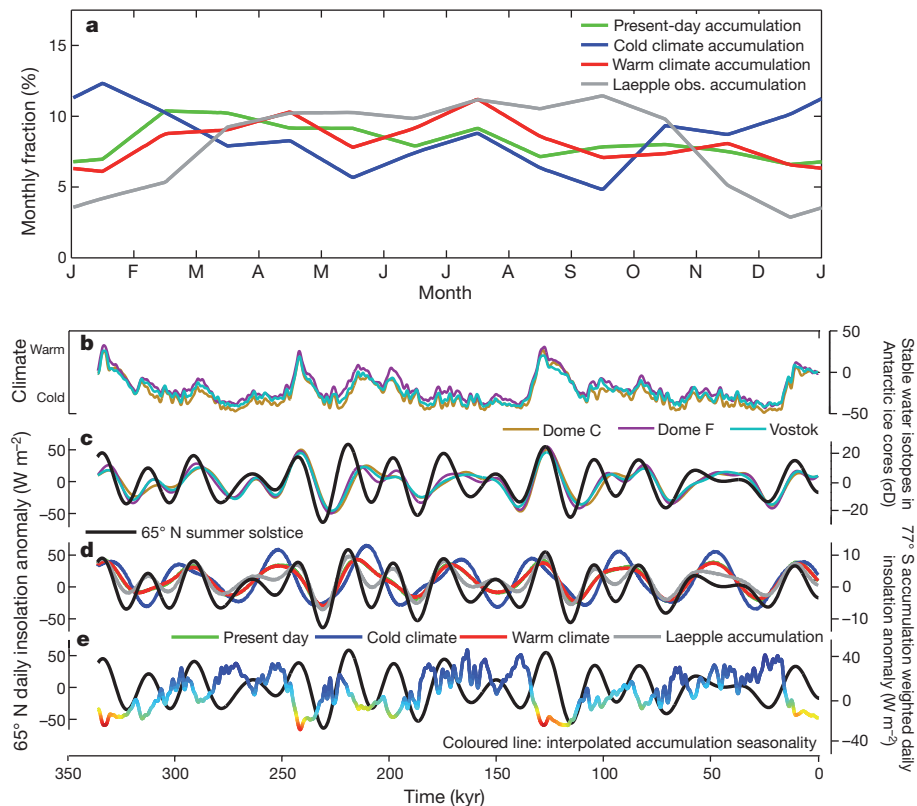


Figure 1 | Accumulation seasonality and the effect of insolation weighting. **a**, Seasonality of accumulation from general circulation model simulations (200 km mean around Dome C site⁵) and the mean of the Laepple *et al.*² East Antarctic Plateau observation set (obs.). Letters on x-axis indicate months of the year (January–December). **b**, Climatic (left axis) East Antarctic Plateau ice-core records of stable water isotopes (right axis), shown on the EDC3 timescale³. An approach we previously described⁴ is used to ensure that differences in core sampling do not affect the results. **c**, The ice-core stable isotope records band-pass filtered between 15 and 50 kyr (right axis). The maximum summer insolation intensity at 65° N (left axis). **d**, Accumulation-weighted insolation at 77° S, using the four estimates of seasonality shown in

a (right axis, same colours as **a**). The maximum summer insolation intensity at 65° N (left axis). **e**, Accumulation-weighted insolation at 77° S, using climate-driven (record shown in **b**) seasonality changes. The accumulation seasonality changes are derived from a linear interpolation of the modelled accumulation seasonality shown in **a**. The colour scheme indicates whether the interpolated accumulation seasonality reflects a warmer or colder climate. The maximum summer insolation intensity at 65° N (left axis). The lack of agreement between the two lines suggests that allowing the accumulation seasonality to vary with climate can result in a break down in the relationship between weighted insolation at 77° S and maximum summer insolation intensity at 65° N.

The different climate simulations show different accumulation seasonality. Using the approach of Laepple *et al.*², the impact of these changes in seasonality on the 77° S insolation weighting can be calculated (Fig. 1d). Correlating the 77° S accumulation-weighted insolation values against the 65° N summer intensity (Fig. 1c) yields an *R* value of 0.63 for the present-day simulation; 0.20 for the glacial simulation; and 0.61 for the warm simulation. This range indicates that the ability of the accumulation weighting to match the 65° N record is dependent on accumulation seasonality. Because the climate for the past 340 kyr has mostly been colder than the present day, it could be more representative to use the cold simulation seasonality. In this case, a relatively small amount of the Antarctic ice-core precession and obliquity signal can be explained by seasonal accumulation weighting of Antarctic insolation.

It is more likely that the precipitation and accumulation seasonality is not stable, and that seasonality varied between different past climates. By interpolating between these three initial model estimates of accumulation seasonality, according to past climate (the isotope record shown in Fig. 1b), seasonality can be allowed to vary directly with climate. Weighting the 77° S local insolation according to this 'unstable' seasonality allows a preliminary assessment of the impact of past-climate-driven accumulation changes. The calculation yields a different time series (Fig. 1e, *R* = −0.25). This result indicates that it is possible that past changes in precipitation (and accumulation) could dominate an accumulation-weighted insolation time series. For this reason some further work, examining the idea that the seasonal precipitation cycle is stable through time, is probably required.

METHODS SUMMARY

Long Antarctic stable water isotope records are available for three ice cores^{1,6,7}. These initial climate simulations are run using the HadAM3 atmospheric model. Mean monthly Dome C results are obtained using a 200-km averaging radius⁸. A colder than present-day simulation is based on observed last glacial maximum

sea-surface temperatures^{8,9}. A warmer simulation is based on the Special Report on Emissions Scenarios (SRES) A1B future scenario. Refs 4 and 5 contain further model and simulation details. T. Laepple kindly provided the 'Laepple' Fig. 1a mean accumulation observation time series. Thanks also to I. Eisenman and P. Huybers for providing insolation calculation tools. Unstable orbital-sublimation and orbital-precipitation effects could be investigated using a larger set of palaeoclimate snapshot simulations.

Louise C. Sime¹ & Eric W. Wolff¹

¹British Antarctic Survey, Cambridge CB3 0ET, UK.

e-mail: lsim@bas.ac.uk

1. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles over the past 360,000 years implied by accurately dated Antarctic ice cores. *Nature* **448**, 912–916 (2007).
2. Laepple, T., Werner, M. & Lohmann, G. Synchronicity of Antarctic temperatures and local solar insolation on orbital timescales. *Nature* **471**, 91–94 (2011).
3. Parrenin, F. *et al.* The EDC3 chronology for the EPICA Dome C ice core. *Clim. Past* **3**, 485–497 (2007).
4. Sime, L. C., Wolff, E. W., Oliver, K. I. C. & Tindall, J. C. Evidence for warmer interglacials in East Antarctic ice cores. *Nature* **462**, 342–345 (2009).
5. Sime, L. C. *et al.* Antarctic isotopic thermometer during a CO₂ forced warming event. *J. Geophys. Res.* **113**, D24119 (2008).
6. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the last 800,000 years. *Science* **317**, 793–796 (2007).
7. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
8. Paul, A. & Schäfer-Neth, C. *Gridded Global LGM SST and Salinity Reconstruction. IGBP PAGES/World Data Center for Paleoclimatology, Boulder Data Contribution Series 2003–046* (NOAA/NGDC Paleoclimatology Program, 2003).
9. Paul, A. & Schäfer-Neth, C. Modeling the water masses of the Atlantic Ocean at the Last Glacial Maximum. *Paleoceanography* **18**, 1058 (2003).

Author Contributions L.C.S. conducted the original model simulations, carried out the insolation weighting calculations, and co-wrote the manuscript. E.W.W. motivated the work and co-wrote the paper.

Competing financial interests: declared none.

doi:10.1038/nature10613

Laepple *et al.* reply

REPLYING TO L. C. Sime & E. W. Wolff *Nature* **479**, doi:10.1038/nature10613 (2011).

Sime and Wolff¹ discuss our local hypothesis for the orbital variability in Antarctic ice cores² by analysing the stability of Antarctic accumulation as simulated by the HadAM3 atmosphere general circulation model (GCM). They propose that a shift of the accumulation maximum towards austral summer in colder climates could dominate the weighted insolation and therefore mask the relationship of the accumulation-weighted signal and Northern Hemisphere summer insolation. We show that the applied model approach is inappropriate to evaluate the accumulation seasonality as current global GCMs are not skilful at modelling present and past accumulation seasonality at the core sites. Furthermore, time variability in accumulation seasonality can weaken or strengthen the accumulation-weighting hypothesis, depending on the time evolution of the seasonality changes.

Sime and Wolff's accumulation estimates from the HadAM3 model exhibit a significantly weaker present-day seasonality than the observational estimates (summer (December–February) – winter (June–August) difference in monthly accumulation fraction; simulated, 0.2%; observed, 6.2% ± 2.0% (two standard deviations)) and show no significant correlation to the observational estimate (*R* = 0.33, *P* > 0.3), (Fig. 1a). As a consequence, the insolation weighted by the modern accumulation cycle of Sime and Wolff¹ only shows a weak

precession variability and a weaker relationship to Northern Hemisphere summer insolation (*R* = 0.63) compared to the insolation weighted by the observed accumulation estimate² (*R* = 0.96).

These differences between the observational and HadAM3-model-based estimates of accumulation seasonality make it worthwhile to test the model skill by comparing different GCMs. Analysing the modern estimate of accumulation seasonality from 22 GCMs from the World Climate Research Programme's Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset we detected no consistent pattern: 13 models simulate less accumulation in austral summer (November–February (NDJF)) than average; 9 simulate the opposite pattern (Fig. 1a). Looking at a subset of ten models, which have the highest skill in modelling Antarctic climate³, does not improve the consistency (four models show a NDJF maximum, six a minimum).

This large spread in modern accumulation seasonality estimates shows that it is unlikely that current GCMs are skilful in modelling climate dependent seasonality changes on the East Antarctic Plateau. Indeed, multimodel studies based on the Paleoclimate Modelling Intercomparison Project GCMs that analyse the change in precipitation⁴ or accumulation seasonality⁵ between the Last Glacial Maximum (LGM) and modern climate found that the amplitude and sign of the

BRIEF COMMUNICATIONS ARISING

changes are model dependent. In the Dome C area, half of the investigated climate models⁵ (five of nine) predicted changes in the accumulation seasonality causing a warm bias for glacial conditions; the other half proposed a cold bias, and the mean effect over all model results was indistinguishable from zero.

Owing to this deficit of global GCMs in simulating Antarctic accumulation seasonality in a correct and consistent manner we decided to refrain from relying on GCM results for present or past accumulation estimates in our study.

However, even if current climate models cannot reliably provide the sign and amplitude of glacial–interglacial changes in accumulation seasonality on the East Antarctic Plateau, it is useful to examine the effect of time-variable accumulation seasonality in a general manner. The simulations from Sime and Wolff¹ propose a significant shift of the accumulation towards austral summer in colder climates. This results in a weighted insolation signal opposing the observed

isotopic record and negatively correlates with Northern Hemisphere summer insolation. In cases of either no change in seasonality², or a shift of accumulation towards winter in colder climates, the weighted insolation signal positively correlates with the isotopic signal. Any accumulation shift towards winter in colder climates strengthens the orbital signal in phase with the Northern Hemisphere summer insolation (Fig. 1b).

The three climate simulations used by Sime and Wolff¹ (LGM, modern and IPCC-AR4 emissions scenario A1B) simulate accumulation changes in response to different sea-surface temperature and greenhouse-gas boundary conditions. Although they might cover the temperature conditions of the last 350 kyr they do not represent the range of insolation conditions of this time period. There is no reason to believe that the considered effects have a stronger influence on the accumulation seasonality than the strong seasonal variations of the local radiative forcing. Combining the estimated insolation-driven change of accumulation seasonality (supplementary note 3 in

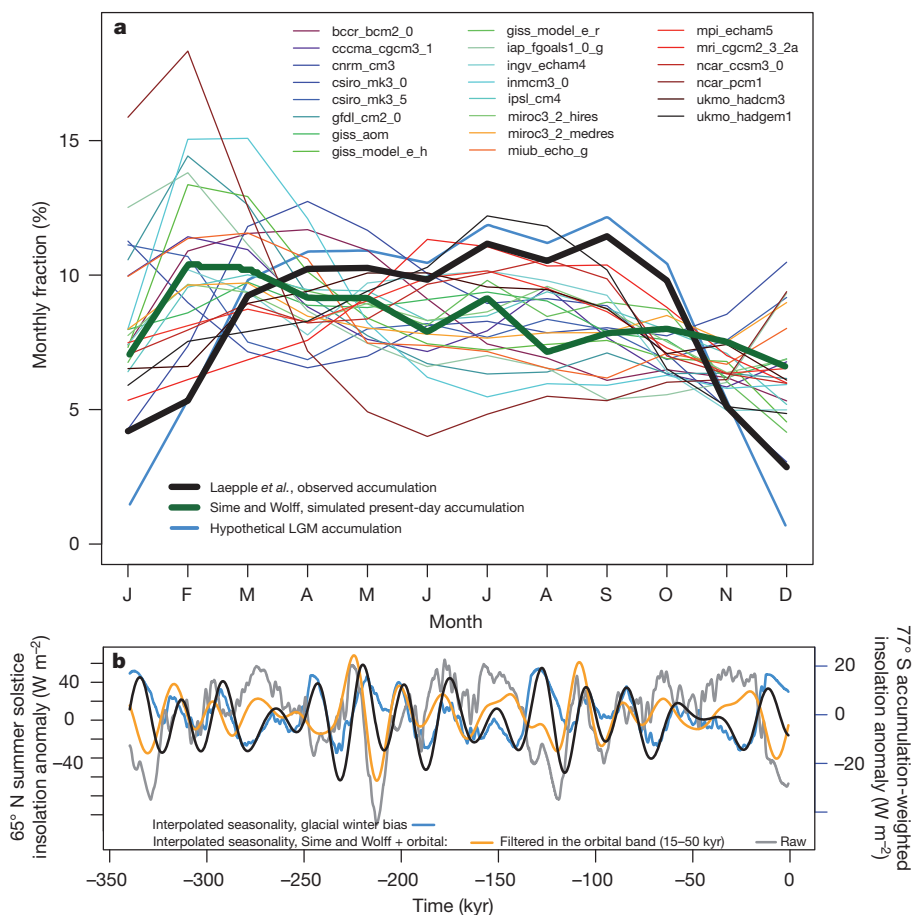


Figure 1 | Model dependence of accumulation seasonality and effects of seasonality instability. **a**, Simulated and observed modern climatological seasonal cycle of Antarctic accumulation (200 km around Dome C site). The stacked observation set² (black) and the simulated present-day accumulation of Sime and Wolff¹ (green) are statistically unrelated. The seasonal cycle, modelled by the CMIP3 climate models is strongly model dependent (thin coloured lines). An arbitrarily chosen glacial accumulation seasonality in the range of the multimodel results at Dome C⁵ (cold glacial bias of $\sim 1.5^\circ\text{C}$) is shown as a bold blue line. Letters on x-axis indicate months of the year (January–December). **b**, Accumulation-weighted insolation at 77°S (right axis) using a climate-driven seasonality as in Sime and Wolff¹, but with the observed accumulation (a, black) as present day and the glacial accumulation (a, blue) as cold climate end member. This seasonality, which varies with the climate, leads to an orbital

signal with increased amplitude compared to the constant accumulation² and correlated to Northern Hemisphere summer insolation (black, left axis). Accumulation-weighted insolation, using the climate-variable simulated seasonality from Sime and Wolff¹, but also accounting for insolation-dependent sublimation results in a signal with precession variability in phase with Northern Hemisphere summer insolation (orange). Sublimation sensitivity is included as described in supplementary note 3 in ref. 2 using the climate-model-derived sublimation sensitivity of $b = 0.8\%$ per W m^{-2} . Under fixed insolation conditions, the sublimation amount is assumed to be the modern observed fraction² (14%) of the accumulation amount. For the modern insolation values, the accumulation seasonalities from Sime and Wolff¹ are used. L. Sime kindly provided the simulated accumulation seasonalities and the weights for the simulated climate-driven seasonality.

ref. 2), with either of the climate dependencies of accumulation seasonality discussed earlier, results in an orbital signal in the weighted insolation correlated to Northern Hemisphere summer insolation in line with our hypothesis. Even in the variable accumulation case of Sime and Wolff¹, additionally accounting for the insolation-dependent accumulation changes the correlation in the precession band (15–30 kyr), the determining factor for the local hypothesis, from $R = -0.54$ to $R = 0.60$.

There is no consistent skill in GCM modelling of modern accumulation on the East Antarctic Plateau and the sign and amplitude of the past seasonality change are unclear. Our assumption of a constant seasonality based on modern observations² is thus a reasonable hypothesis that also provides a local explanation for the observed orbital signal. Ultimately, we agree with Sime and Wolff¹ that a fuller understanding of the seasonality in climate and accumulation in Antarctica⁶ is needed for a robust interpretation of the orbital signal in the isotope record.

Thomas Laepple¹, Martin Werner¹ & Gerrit Lohman¹

¹Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany.

e-mail: thomas.laepple@awi.de

1. Sime, L. C. & Wolff, E. W. Antarctic accumulation seasonality. *Nature* **479**, doi:10.1038/nature10613 (2011).
2. Laepple, T., Werner, M. & Lohmann, G. Synchronicity of Antarctic temperatures and local solar insolation on orbital timescales. *Nature* **471**, 91–94 (2011).
3. Connolley, W. M. & Bracegirdle, T. J. An Antarctic assessment of IPCC AR4 coupled models. *Geophys. Res. Lett.* **34**, L22505 (2007).
4. Masson-Delmotte, V. *et al.* Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints. *Clim. Dyn.* **26**, 513–529 (2006).
5. Krinner, G. & Werner, M. Impact of precipitation seasonality changes on isotopic signals in polar ice cores: a multi-model analysis. *Earth Planet. Sci. Lett.* **216**, 525–538 (2003).
6. Huybers, P. Antarctica's Orbital Beat. *Science* **325**, 1085–1086 (2009).

doi:10.1038/nature10614